Ph.D. DISSERTATION

Channel Estimation and Feedback Techniques for Massive MIMO Systems

Massive MIMO 시스템을 위한 채널 추정 및 피드백 기법

Yonghee Han

February 2017

Department of Electrical Engineering and Computer Science
College of Engineering
Seoul National University

Ph.D. DISSERTATION

Channel Estimation and Feedback Techniques for Massive MIMO Systems

Massive MIMO 시스템을 위한 채널 추정 및 피드백 기법

Yonghee Han

February 2017

Department of Electrical Engineering and Computer Science
College of Engineering
Seoul National University

Channel Estimation and Feedback Techniques for Massive MIMO Systems

Advisor: Jungwoo Lee

Presented to the Graduate School of Seoul National University
in Partial Fulfillment of the Requirements for

The Degree of Doctor of Philosophy

November 2016

by

Yonghee Han
Department of Electrical Engineering and Computer Science
College of Engineering
Seoul National University

This dissertation is approved for

The Degree of Doctor of Philosophy

December 2016

Chairman _____
Vice chairman _____
Member _____
Member _____
Member _____

f Channel Estimation and Feedback Techniques for Massive MIMO Systems

Massive MIMO 시스템을 위한 채널 추정 및 피드백 기법

지도교수 이 정 우
이 논문을 공학박사 학위논문으로 제출함

2016년 11월

서울대학교 대학원

전기·컴퓨터 공학부

한 용 희

한용희의 공학박사 학위 논문을 인준함

2016년 12월

위 원 장: _____
부위원장: _____
위 원: _____
위 원: _____
위 원: _____

Abstract

To meet the demand of high throughput in next generation wireless systems, various directions for physical layer evolution are being explored. Massive multiple-input multiple-output (MIMO) systems, characterized by a large number of antennas at the transmitter, are expected to become a key enabler for spectral efficiency improvement. In massive MIMO systems, thanks to the orthogonality between different users' channels, high spectral and energy efficiency can be achieved through simple signal processing techniques. However, to get such advantages, accurate channel state information (CSI) needs to be available, and acquiring CSI in massive MIMO systems is challenging due to the increased channel dimension. In frequency division duplexing (FDD) systems, where CSI at the transmitter is achieved through downlink training and uplink feedback, the overhead for the training and feedback increases proportionally to the number of antennas, and the resource for data transmission becomes scarce in massive MIMO systems. In time division duplexing (TDD) systems, where the channel reciprocity holds and the downlink CSI can be obtained through uplink training, pilot contamination due to correlated pilots becomes a performance bottleneck when the number of antennas increases.

In this dissertation, I propose efficient CSI acquisition techniques for various massive MIMO systems. First, I develop a downlink training technique for FDD massive MIMO systems, which estimates the downlink channel with small overhead. To this end, compressed sensing tools are utilized, and the training overhead can be highly reduced by exploiting the previous channel information. Next, a limited feedback scheme is developed for FDD massive MIMO systems. The proposed scheme reduces the feedback overhead using a dimension reduction technique that exploits spatial and

temporal correlation of the channel. Lastly, I analyze the effect of pilot contamination, which has been regarded as a performance bottleneck in multi-cell massive MIMO systems, and propose two uplink training strategies. An iterative pilot design scheme is developed for small networks, and a scalable training framework is also proposed for networks with many cells.

keywords: massive MIMO, large-scale antenna systems, channel estimation, limited feedback, pilot contamination

student number: 2011-20954

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Massive MIMO

To meet the exploding demand for communication over the wireless medium, next generation wireless systems (e.g. 5G) are required to greatly improve throughput, reliability, and latency compared to their predecessors, and various directions for the wireless evolution are being explored [1, 2]. Among them, massive multiple-input multiple-output (MIMO) is a potential enabler for the throughput improvement. Over the past two decades, MIMO systems have been thoroughly investigated to manifest their advantage such as multiplexing and diversity gains, and contributed to improving the throughput and reliability of wireless systems such as cellular networks and wireless local area networks (WLAN) [3]. Recently, it was shown that the spectral and energy efficiency of MIMO systems can be unprecedentedly increased by increasing the number of antennas [4, 5], and massive MIMO systems, also known as large-scale antenna systems, large-scale MIMO, very large MIMO, etc., are drawing a interests as a key enabler for high-throughput future wireless systems.

In [5], potential of massive MIMO was firstly revealed by analyzing the asymp-

totic spectral efficiency of single-user and multi-user MIMO systems. The seminal analysis, which relies on the law of large numbers, implies that the spectral efficiency of MIMO systems can be increased without a bound by simply increasing the number of antennas, while maintaining a constant transmit power, and such benefit can be reaped through simple signal processing techniques such as maximal ratio transmission/combining (MRT/MRC). Due to such advantages, massive MIMO has drawn significant research interests, and various aspects of massive MIMO systems have been investigated [6–8]. In [6], with the mathematical tool of deterministic equivalents, the analysis of [5] was extended to more realistic settings where the number of antennas increases proportionally to the number of users. The performance of two linear precoders (MRT and zero-forcing (ZF)) were compared in [7], and the preferable precoder was investigated in various operation scenarios. In [8], the benefit of massive MIMO systems was investigated in terms of spectral efficiency, and it was shown that the transmit power can be reduced inversely proportional to the number of antennas given the perfect channel state information (CSI).

Along with the promising theoretical results, practical challenges for the realization of the massive MIMO benefits have also been posed [9]. For example, performance degradation due to hardware impairments such as amplifier non-linearities, I/Q-imbalance and phase noise was quantitatively studied [10]. The impact of cost-efficient hardware such as constant envelope precoders [11] or one-bit ADCs [12] was also investigated. Implementation issues such as low-complexity detection algorithms were also addressed to realize practical massive MIMO systems [13]. Along with these practical issues, acquiring CSI poses a fundamental challenge in massive MIMO systems due to the large dimension of channels, which is addressed in this dissertation.

1.2 CSI Acquisition in Massive MIMO Systems

Since the initial analysis of [5], the CSI acquisition has been pointed as a fundamental challenge in massive MIMO systems. The method of channel acquisition in MIMO systems depends on the operation mode and the communication direction.

In frequency division duplexing (FDD) systems, where the uplink and downlink transmissions are in separated frequency bands, the base station (BS) acquires the CSI for downlink transmission from the uplink feedback, while users estimate their corresponding channels to feed back from the downlnk training. Since conventional schemes for the downlink training and feedback demand overheads proportional to the number of BS antennas, the early studies of massive MIMO assumes time division duplexing (TDD), where the uplink-downlink reciprocity holds and the required length of the uplink training is not proportional to the number of downlink transmit antennas. However, since most of the contemporary cellular networks adopt FDD, developing efficient training and feedback schemes for FDD massive MIMO is of great importance for designing backward compatible 5G networks.

Conventional channel estimation schemes adopted for small-scale MIMO system may consume prohibitive downlink resource in massive MIMO systems. Some recent work has been devoted to estimating a temporally/spatially correlated massive MIMO channel with a short training sequence exploiting the correlated nature of the massive MIMO channel [19, 20]. In [19], Kalman filter-based channel estimation was studied and a pilot design minimizing the mean squared error (MSE) was developed. With this approach, the channel can be estimated using a short training sequence given transmitter knowledge of previous channel outputs received during training or channel statistics. Feedback-assisted downlink training techniques that are performed without the transmitter's knowledge of channel statistics were proposed in [20].

Compressed sensing (CS) is an attractive strategy to estimate a sparse channel with

a short sequence. According to the CS theory, a sparse vector can be recovered from a compressive measurement [21–24], and a sparse channel can be estimated efficiently leveraging the CS theory [25, 26]. Since massive MIMO channels exhibit limited scatters compared to the large number of antennas, it can be represented by a sparse vector in the virtual angular domain [27, 28]. Exploiting the sparsity, CS has been utilized for the estimation [29, 30] of massive MIMO channels. In [29], channel estimation via low-rank approximation was proposed for TDD multiuser massive MIMO systems. In [30], a CS-based channel estimation method was proposed, which exploits the joint sparsity of channel. It is worth mentioning that since millimeter wave channels tend to have only few significant paths, and directional beamforming with a large number of antennas is essentially required to mitigate the severe attenuation [32], CS-based channel estimation is more effective for millimeter wave systems [33–35].

Since accuracy of CSI at the transmitter (CSIT) has a critical impact on the performance of multi-user MIMO systems, extensive studies have been devoted to limited feedback in conventional (small-scale) MIMO systems [57–60, 62–66]. For spatially uncorrelated channels, Grassmannian codebook was designed to maximize the minimum chordal distance between codewords [58], and performance of random vector quantization was analyzed in single-user [59] and multi-user MIMO systems [60]. In [61], a hierarchical codebook for multi-user MIMO systems was also developed, along with a corresponding scheduling scheme. In [62] and [63], codebooks for spatially correlated channels were developed by rotating an independent and identically distributed (i.i.d.) codebook to align with the channel distribution. For temporally correlated channels, differential feedback concept has been investigated, where a codebook is generated around the previous quantized vector [64–66].

To achieve the rate of the perfect CSIT within a constant gap with such schemes, the codebook size (in bits) needs to scale linearly with the number of the transmit an-

tennas [60]. Hence, developing a CSI feedback strategy that provides accurate CSIT with a low overhead and a low encoding complexity has more important meaning in massive MIMO systems. For massive MIMO systems, a trellis code-based quantization scheme was developed capturing the duality between source encoding (quantization) and channel decoding in [15] and extended to spatially and temporally correlated channel in [67]. As another attractive way, dimension reduction of feedback was studied in [18, 31, 68–71], which is based on the observation that spatially correlated massive MIMO channels have degrees of freedom much smaller than the dimension of channels. In [68], grouping of highly correlated antenna elements was proposed to reduce the dimension of feedback. In [31] and [69], a compressive sensing scheme with random projection was applied to feedback. In [18] and [70], a new framework for FDD massive MIMO, referred to as joint spatial division and multiplexing (JSDM), was proposed under the assumption of clustered users. In JSDM, users are divided into groups such that the users of each group have similar covariance matrix, and channels are projected into the nullspace of other group channel spaces to reduce the dimension of the feedback information. A joint design of training pilot and feedback strategy was proposed for multi-user systems exploiting the spatial channel correlation in [71].

In MIMO systems operating with TDD, the required resource for uplink training is proportional to the number of users, not antennas, which is favorable for massive MIMO systems. However, since the resource available for the training is still finite, and the uplink pilot can be correlated in multi-cell systems. Pilot contamination caused by the correlated pilot has been pointed out as the performance bottleneck. It is worth noting that the pilot contamination is observed in the uplink transmission of FDD systems. Due to the importance of pilot contamination, various mitigation techniques have been developed. In [84], the use of time-shifted pilot sequences was proposed to avoid pilot contamination. In [85], a pilot assignment strategy that assigns the same pilot

sequence to UEs with approximately orthogonal channel covariances was proposed. A practical approach that assigns orthogonal sequences to neighboring cells was investigated in [86]. The scheme, referred to as pilot reuse, has been modified in various ways. Fractional pilot reuse, which assigns orthogonal sequence only to UEs in the cell edge, was proposed in [87], and an adaptive pilot assignment scheme was proposed for non-symmetric networks in [88]. The optimal reuse factor for the pilot reuse was investigated in [90]. The use of non-orthogonal pilot was proposed in [89] under the assumption of one user per cell and a symmetric network topology.

1.3 Contributions and Organization

In this dissertation, challenges in acquiring CSI in massive MIMO systems are addressed with various system models. For FDD systems, an efficient downlink training scheme and a novel CSI feedback framework are proposed. For TDD systems and the uplink transmission of FDD systems, pilot design strategies alleviating the pilot contamination are proposed.

In Chapter 2, a CS-aided channel training scheme is developed for the downlink of FDD massive MIMO systems, which exploits the observation that the channel statistics change slowly in time. Unlike the previous work, which directly applies CS to estimate the angular domain channel, the proposed scheme utilizes a conventional least squares approach and a CS technique simultaneously exploiting partial support information. Simulation results show that the proposed scheme can estimate the channel with a reduced pilot overhead even when conventional CS cannot be applied.

In Chapter 3, a projection-based differential feedback (PBDF) framework is proposed for FDD massive MIMO systems. To reduce the feedback overhead exploiting the spatial and temporal correlation simultaneously, in the proposed framework, a difference between original and predicted vectors is projected, and quantization is

performed in a smaller dimensional subspace. An appropriate projection for the proposed framework is also derived. Simulation results show that the proposed scheme achieves a large portion of potential throughput gain of massive MIMO systems with small amount of channel feedback.

In Chapter 4, the effect of the pilot contamination is analyzed, and two uplink pilot design strategies for alleviating the pilot contamination are proposed for multi-cell massive MIMO systems. First, a pilot design algorithm based on alternating minimization is developed. Second, a scalable channel training framework, referred to as generalized pilot reuse scheme, is presented, and it is shown that the Grassmannian subspace packing provides an optimal sequence set for the framework. Contrary to existing pilot reuse protocols that restrict the pilot length to be an integer multiple of the number of users, the proposed methods can design a pilot with an arbitrary length, and simulation results show that the proposed pilot design strategies provide an opportunity to achieve a higher throughput.

1.4 Notations

A vector (matrix) is written in boldface as $\mathbf{a}$ ($\mathbf{A}$). $\mathbf{a}^*$ ($\mathbf{A}^*$), $\mathbf{a}^T$ ($\mathbf{A}^T$), and $\mathbf{a}^H$ ($\mathbf{A}^H$) denote the complex conjugate, the transpose, and the conjugate transpose of $\mathbf{a}$ ($\mathbf{A}$), respectively. For a length-n vector $\mathbf{a}$ and an integer $p \geq 1$, $\|\mathbf{a}\|_p$ denotes the p-norm of $\mathbf{a}$, i.e., $\|\mathbf{a}\|_p = (\sum_{i=1}^{n} |a_i|^p)^{1/p}$, where a_i is the ith component of $\mathbf{a}$, and $|a|$ denotes the absolute value of scalar a. The 0-norm of $\mathbf{a}$ $\|\mathbf{a}\|_0$ is defined as the number of nonzero elements in $\mathbf{a}$. $\mathsf{tr}(\mathbf{A})$ denotes the trace of $\mathbf{A}$. $\mathbf{A}^{-1}$ and $\mathbf{A}^+ = \mathbf{A}^H(\mathbf{A}\mathbf{A}^H)^{-1}$ are the inverse matrix and the pseudo-inverse matrix of $\mathbf{A}$, respectively. $\mathrm{diag}(\mathbf{a})$ denotes a diagonal matrix whose diagonal is $\mathbf{a}$. $\mathrm{Re}(\mathbf{a})$ and $\mathrm{Im}(\mathbf{a})$ denote the real and imaginary parts of $\mathbf{a}$. $\mathrm{sign}(\mathbf{a})$ is the function that returns the sign of $\mathbf{a}$. $\mathbf{I}_n$ is the $n \times n$ identity matrix. $(\mathbf{A})_{\mathcal{I}}$ ($(\mathbf{a})_{\mathcal{I}}$) is a submatrix (subvector) formed by collecting the columns

(elements) of $\mathbf{A}$ ($(\mathbf{a})$) which correspond to indices in the set $\mathcal{I}$, and $(\mathbf{A})_{\mathcal{I}}^{T}$ ($(\mathbf{A})_{\mathcal{I}}^{H}$) denotes its transpose (conjugate transpose). For set $\mathcal{S}$, $|\mathcal{S}|$ denotes the cardinality of $\mathcal{S}$, and $\mathcal{S}^{c}$ is the complement of set $\mathcal{S}$. $\mathbb{E}[\cdot]$ stands for the expectation operator. $\mathcal{CN}(\mu, \mathbf{\Sigma})$ represents a complex Gaussian distribution with a mean μ and covariance matrix $\mathbf{\Sigma}$.

Chapter 2

Compressed Sensing-Aided Downlink Training

There is much discussion in industry and academia about possible technical solutions to address the growth in demand for wireless broadband. Massive multiple-input multiple-output (MIMO) systems are one of the most popular solutions to addressing this broadband demand in fifth generation (5G) cellular systems. Massive MIMO systems employ 10s or 100s of antennas at the base station to enable advanced multiuser MIMO communications. To reap the massive MIMO throughput gain, coherent transmission exploiting accurate channel state information at the transmitter (CSIT) is required. While it is expected that many 5G systems will employ frequency division duplexing (FDD), channel sounding for FDD systems requires a large pilot overhead, which usually scales proportionally to the number of transmit antennas. To resolve this problem, a compressed sensing (CS)-aided channel estimation scheme is proposed, which exploits the observation that the channel statistics change slowly in time. By utilizing a conventional least squares (LS) approach and a CS technique simultaneously, the proposed scheme reduces the pilot overhead. Simulation results show that the proposed scheme can estimate the channel with a reduced pilot overhead even when conventional CS cannot be applied.

2.1 Introduction

To meet the demand for high throughput in next generation (e.g. 5G) cellular networks, various directions for physical layer evolution are being explored [1]. Among them, massive multiple-input multiple-output (MIMO) systems, characterized by a large number of antennas at the transmitter, are drawing significant interest for potential standardization.

Since the seminal analysis in [5], work has shown that massive MIMO systems can provide high throughput and energy efficiency improvements with simple transmission/reception techniques [4, 5]. However, such advantages are based upon the premise of accurate channel state information (CSI) at the transmitter (CSIT). Due to the large dimension of the channels, CSIT acquisition is one of the most challenging problems in massive MIMO systems. To exploit the channel reciprocity and acquire CSIT with an uplink training sequence whose length is not proportional to the number of downlink transmit antennas, most massive MIMO research has assumed two-way communication is facilitated through time division duplexing (TDD) [4–7]. However, the complicated calibration required for TDD reciprocity and the adoption of frequency division duplexing (FDD) in most of the contemporary cellular networks make FDD an attractive option for 5G networks.

Accordingly, CSIT acquisition in FDD massive MIMO systems is of great importance. In FDD systems, CSIT can be achieved through downlink training and feedback from the receiver. Due to the large dimension of the channel, both the channel training and feedback are challenging in massive MIMO systems. On the feedback issue, some techniques have been proposed to reduce the feedback overhead and encoding complexity by leveraging structured quantizers such as trellis-coded quantization [14–16] or projecting the channel into a lower dimensional subspace [17, 18]. CSIT feedback, however, hinges on accurate downlink channel estimation, which is the focus of this

chapter.

Conventional channel estimation schemes adopted for small-scale MIMO systems, such as least squares (LS)-based schemes, require the length of the training sequence to be proportional to the number of transmit antennas and become inefficient in massive MIMO systems. To tackle this problem, some recent work has been devoted to estimating a massive MIMO channel with a short training sequence exploiting the correlated nature of the massive MIMO channel [19, 20]. In [19], Kalman filter-based channel estimation was studied and a pilot design minimizing the mean squared error (MSE) was developed. With this approach, the channel can be estimated using a short training sequence given transmitter knowledge of the previous channel outputs received during training or channel statistics. However, acquiring the channel statistics for non-stationary massive MIMO channels requires additional feedback. While training techniques proposed in [20] can be performed without the transmitter's knowledge of channel statistics, these schemes are practically effective only for slow-fading channels with high temporal correlation since they are based on adaptive filtering.

Compressed sensing (CS) is an attractive strategy to estimate a sparse channel with a short sequence when the channel statistics are unknown. According to the CS theory, a sparse vector can be recovered from a compressive measurement [21–24], and a sparse channel can be estimated efficiently leveraging the CS theory [25, 26]. Since massive MIMO channels exhibit limited scatters compared to the large number of antennas, it can be represented by a sparse vector in the virtual angular domain [27,28]. Exploiting the sparsity, CS has been utilized for the estimation [29], [30] and feedback design [31] for massive MIMO channels. In [29], channel estimation using low-rank approximation was proposed for TDD multiuser massive MIMO systems. In [30], a CS-based channel estimation method was proposed, which exploits the joint sparsity of the channel. However, considering that about $\mathcal{O}(k \log(M))$ measurements are required

for the recovery of a k-sparse length-M vector [21], direct application of CS gives very little training overhead reduction with the sparsity levels prevalent in non-millimeter (centimeter or larger) wave massive MIMO channels. It is worth mentioning that since millimeter wave channels tend to have only a few significant paths and directional beamforming with a large number of antennas is essentially required to mitigate the severe attenuation [32], CS-based channel estimation is more suitable for millimeter wave systems [33–35].

In this chapter, I propose a novel channel estimation scheme for FDD massive MIMO systems, which combines LS and CS techniques. A key observation for the proposed scheme is that even though the channel statistics change with time, the rate of change is much lower than that of the channel response. Since the variation in channel statistics occurs with this appearance and disappearance of scatterers at specific locations, the channel space intuitively evolves slowly if parameterized correctly. This approach has previously been modeled as a birth-death evolution of scattering clusters in [36]. From a CS viewpoint, this means that the support of the previous channel can be a good predictor for the support of the current channel. In other words, only a small number of elements are nonzero outside the support of the previous channel. Hence, it is possible to separate the channel vectors into a dense vector (by projecting into the previous support) and a sparse vector (by projecting into the null space of the previous support). The two vectors are estimated with different strategies. The LS and CS techniques are used to estimate the dense and sparse vectors, respectively. With the separation, the proposed scheme can reduce pilot overhead and improve tracking of the channel subspace.

The rest of this chapter is organized as follows. In Section 2.2, the channel model and the downlink channel estimation problem are introduced. In Section 2.3, the CS-aided channel estimation scheme is proposed for FDD massive MIMO systems, and

some practical issues are discussed in Section 2.4. In Section 2.5, numerical results are provided. Finally, conclusion is drawn in Section 2.6.

2.2 System Model

A single-user massive MIMO system with M ($M \gg 1$) transmit antennas and a single receive antenna is considered. The received signal at the nth symbol time is given by

$$y[n] = \sqrt{\rho}\mathbf{h}^H[n]\mathbf{x}[n] + z[n], \tag{2.1}$$

where $\mathbf{h}[n] \in \mathbb{C}^{M\times 1}$ is the channel vector at the nth symbol time, $\mathbf{x}[n] \in \mathbb{C}^{M\times 1}$ is the transmit signal vector at the nth symbol time with $\mathbb{E}[\|\mathbf{x}[n]\|_2^2] = 1$, ρ is the signal-to-noise ratio (SNR), and $z[n] \in \mathbb{C}$ is the additive white noise at the nth symbol time with the distribution of $\mathcal{CN}(0, 1)$.

2.2.1 Channel Model

A block-fading model is assumed, where the channel remains constant within a fading block of L consecutive channel uses and channels of different blocks are uncorrelated. The fading model can be expressed as

$$\mathbf{h}_i = \mathbf{h}\left[(i-1)L + l\right], \quad l = 1, 2, \cdots, L, \tag{2.2}$$

where $\mathbf{h}_i$ denotes the channel vector of the ith block and

$$\mathbb{E}\left[\mathbf{h}_i^H \mathbf{h}_j\right] = 0, \qquad i \neq j. \tag{2.3}$$

The channel vector can be represented in the angular domain with a proper transformation. Let $\mathbf{s}_i \in \mathbb{C}^{M\times 1}$ be the angular domain representation of channel vector $\mathbf{h}_i$. Then, the channel can be written as

$$\mathbf{h}_i = \mathbf{\Psi}\mathbf{s}_i, \tag{2.4}$$

where $\mathbf{\Psi} \in \mathbb{C}^{M \times M}$ is a unitary matrix representing the angular domain transformation [27, 37]. Since the transformation matrix is determined by the structure of the antenna array, I assume that $\mathbf{\Psi}$ is fixed and known. If the antennas are placed in a uniform linear array (ULA), for example, the transformation is given as

$$\mathbf{\Psi}_{\text{ULA}} = \mathbf{F}_M, \tag{2.5}$$

where $\mathbf{F}_M$ is the $M \times M$ discrete Fourier transform (DFT) matrix whose elements are $[\mathbf{F}_M]_{m,n} = \frac{e^{-j2\pi mn/M}}{\sqrt{M}}$. If the antennas form an $M_v \times M_h$ uniform planar array (UPA), the two-dimensional DFT (2D-DFT) can be used for the angular domain representation, i.e.,

$$\mathbf{\Psi}_{\text{UPA}} = \mathbf{F}_{M_v} \otimes \mathbf{F}_{M_h}, \tag{2.6}$$

where $\otimes$ denotes the Kronecker product.

Due to the compact deployment of antennas and limited scattering environment, massive MIMO channels tend to be spatially correlated and have fewer degrees of freedom than the number of antennas [4]. In the angular domain, this phenomenon is manifested as a substantial number of elements in $\mathbf{s}_i$ whose values are either zero or close to zero. Considering this fact, I assume that $\mathbf{s}_i$ has at most $k < M$ effective elements and the other elements have negligible values. In other words, a linear combination of columns of $\mathbf{\Psi}$ corresponding to the k largest coefficients can approximate the channel with negligible error. Let $\Omega_i \subset \{1, 2, ..., M\}$ be the support set of $\mathbf{s}_i$, i.e., the set of indices corresponding to the dominant elements, which I assume satisfy $|\Omega_i| = k$. Moreover, I denote the set of indices with negligible values by the complement of the support $\Omega_i^c = \{1, 2, ..., M\} \setminus \Omega_i$. Then the assumption is expressed as

$$\mathbf{h}_i \approx \sum_{j \in \Omega_i} \mathbf{s}_i(j) \boldsymbol{\psi}_j, \tag{2.7}$$

where $\mathbf{s}_i(j)$ is the jth element of the vector $\mathbf{s}_i$ and $\boldsymbol{\psi}_j$ is the jth column of $\mathbf{\Psi}$. In

compressed sensing theory, $\mathbf{s}_i$ is said to be *approximately sparse* if k is much smaller than M, i.e., $k \ll M$ [21].

In practical massive MIMO systems, the support of $\mathbf{s}_i$ is determined by the geometry of the receiver and scattering environment [27, 28]. Reflecting the observation that the environment may change slowly in realistic deployment scenarios and can be modeled as an evolution of scattering clusters [36], it is assumed that the supports of consecutive fading blocks are similar to each other (i.e., Ω_{i-1} and Ω_i share many common entries). Let k_s be the maximum number of support elements that are newly added to Ω_i from those not included in Ω_{i-1}, i.e.,

$$|\Omega_i \setminus \Omega_{i-1}| \leq k_s. \tag{2.8}$$

This means that at most k_s elements are newly added to the support of $\mathbf{s}_i$ from those not included in the support of $\mathbf{s}_{i-1}$. Then, the slow rate of change of the support can be expressed as

$$k_s \ll |\Omega_{i-1}^c|. \tag{2.9}$$

As a result, most of the elements corresponding to the previous support Ω_{i-1} are nonzero while most of the elements corresponding to the complement of previous support Ω_{i-1}^c are zero. In other words, if $\mathbf{s}_i$ is separated into two vectors based on the previous support as

$$\mathbf{s}_{i,d} = (\mathbf{I}_M)_{\Omega_{i-1}}^T \mathbf{s}_i, \tag{2.10}$$

and

$$\mathbf{s}_{i,s} = (\mathbf{I}_M)_{\Omega_{i-1}^c}^T \mathbf{s}_i, \tag{2.11}$$

where $\mathbf{I}_M$ denotes the $M \times M$ identity matrix, then $\mathbf{s}_{i,d} \in \mathbb{C}^k$ is a dense vector, while $\mathbf{s}_{i,s} \in \mathbb{C}^{M-k}$ is a sparse vector. Let $\Lambda_{i,d} \subset \{1, 2, ..., k\}$ and $\Lambda_{i,s} \subset \{1, 2, ..., M-k\}$ be the index sets corresponding to the nonzero elements of $\mathbf{s}_{i,d}$ and $\mathbf{s}_{i,s}$, respectively,

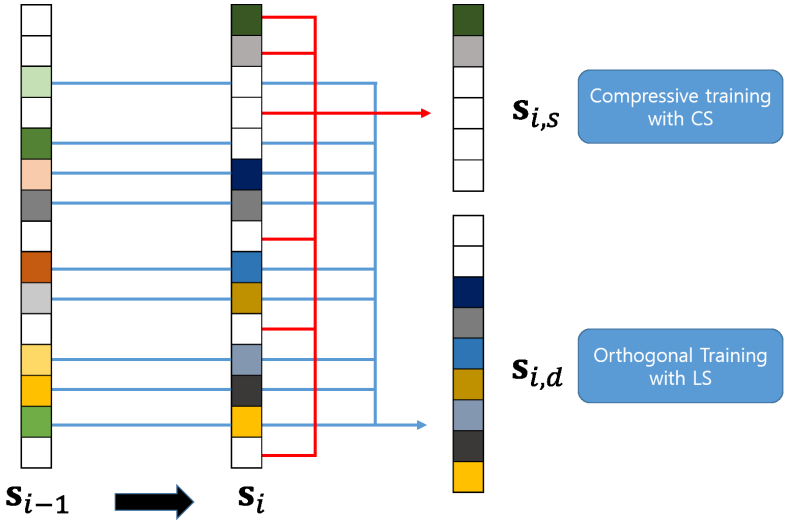

Figure 2.1: Illustration of the channel model and the concept of the proposed channel estimation scheme.

with cardinalities $|\Lambda_{i,d}| \geq k - k_s$ and $|\Lambda_{i,s}| \leq k_s$. Throughout the chapter, I assume that the channel sparsity parameters k and k_s are known. In practical systems, these values need to be chosen with the consideration of long term statistics.

The channel model is depicted in Fig. 2.1, where the white and colored elements of vectors denote the zero and nonzero elements of the vectors, respectively. Although the values of the coefficients (colors) can vary independently, the supports of $\mathbf{s}_{i-1}$ and $\mathbf{s}_i$ have only two different elements. As a result, $\mathbf{s}_i$ can be decomposed into a dense vector $\mathbf{s}_{i,d}$ and a sparse vector $\mathbf{s}_{i,s}$ exploiting the previous support Ω_{i-1}.

2.2.2 Downlink Channel Estimation

To achieve the high spectral efficiency available in massive MIMO systems, coherent transmission exploiting CSIT is required. In FDD systems, CSIT can be achieved through downlink training and uplink feedback. If I neglect the feedback delay, a chan-

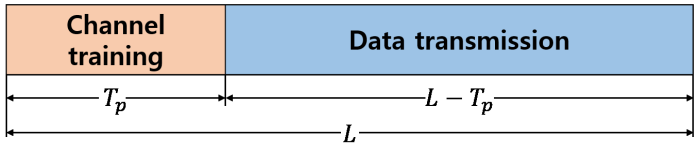

Figure 2.2: Downlink channel block.

nel block can be divided into a channel training period of length T_p and a data transmission period of length $L-T_p$ with a coherence block of length L, as depicted in Fig. 2.2. Since the length of a block is limited by the channel coherence time, the available number of channel uses for downlink transmission decreases as the training time T_p increases. Hence, an efficient channel training scheme that utilizes a small T_p needs to be developed for massive MIMO systems.

The received signal vector during the training period can be expressed as

$$\begin{aligned}\mathbf{y}_{i,T}^{T} &= [y[(i-1)L+1], ..., y[(i-1)L+T_p]] \\ &= \sqrt{\rho_T}\mathbf{h}_i^H \mathbf{X}_T + \mathbf{z}_{i,T}^{T},\end{aligned} \tag{2.12}$$

where $\mathbf{X}_T = [\mathbf{x}_T[1], ..., \mathbf{x}_T[T_p]] \in \mathbb{C}^{M \times T_p}$ is a pilot sequence with the constraint of $\mathsf{tr}(\mathbf{X}_T^H \mathbf{X}_T) = T_p$, ρ_T is the SNR during the training period, and $\mathbf{z}_{i,T}$ is a noise vector. Using the received vector $\mathbf{y}_{i,T}$, the receiver estimates the channel and sends it to the transmitter through a feedback link.

In the absence of knowledge of the channel statistics, the conventional LS approach estimates the channel as

$$\hat{\mathbf{h}}_i^{LS} = \frac{1}{\sqrt{\rho_T}} \left(\mathbf{y}_{i,T}^{T} \mathbf{X}_T^{+}\right)^H, \tag{2.13}$$

where $\mathbf{X}_T^{+}$ is the pseudo-inverse of $\mathbf{X}_T$. However, this approach requires $T_p \geq M$ and causes training to consume a large amount of the downlink resources in massive MIMO systems.

If the transmitter and the receiver know the channel statistics and the support of $\mathbf{s}_i$ a priori, the channel can be estimated using a shorter training sequence. Specifically, a

sequence of length T_p satisfying $T_p \geq k$ is enough to train the coefficients $\mathbf{s}_i(j)$ with $j \in \Omega_i$. The training sequence is constructed as

$$\mathbf{X}_T = \mathbf{\Psi}_{\Omega_i} \tilde{\mathbf{X}}_T, \tag{2.14}$$

where $\mathbf{\Psi}_{\Omega_i}$ is the $M \times k$ matrix obtained by collecting columns of $\mathbf{\Psi}$ corresponding to the indices in Ω_i, and $\tilde{\mathbf{X}}_T \in \mathbb{C}^{k \times T_p}$ is a pilot sequence for training the k nonzero coefficients. Given $T_p \geq k$, a conventional estimation scheme such as LS can estimate the coefficients.

Since the support varies according to the relative location of the receiver and the evolution of the scattering environment, it is difficult to know the exact support of the channel in practical systems. Without knowledge of the support, in general, channel estimation using a pilot of length $T_p < M$ is an underdetermined problem. If $\mathbf{s}_i$ is sparse (or approximately sparse), i.e., $k \ll M$, CS can be utilized for the channel estimation. For the CS approach, the pilot is designed using

$$\mathbf{X}_{CS} = \mathbf{\Psi} \mathbf{\Phi}_{CS}^H, \tag{2.15}$$

where $\mathbf{\Phi}_{CS} \in \mathbb{C}^{T_p \times M}$ is a compressive measurement matrix that satisfies a certain condition. After receiving the training signals, the receiver can estimate the channel using a sparse recovery algorithm such as basis pursuit (BP) or orthogonal matching pursuit (OMP). Much work has been devoted to characterizing the conditions that a measurement matrix $\mathbf{\Phi}_{CS}$ must have to guarantee the successful support recovery in terms of the restricted isometry property (RIP) [23] and mutual coherence [24]. In the literature, it can be found that $T_p = \mathcal{O}(k \log(M/k))$ measurements are required to recover a k-sparse channel, and $T_p \approx 4k$ is usually accepted in many practical applications [21]. Although these results provide great opportunities to reduce the length of training sequence, massive MIMO channels are hardly sparse enough to directly apply CS unless a millimeter wave band is used.

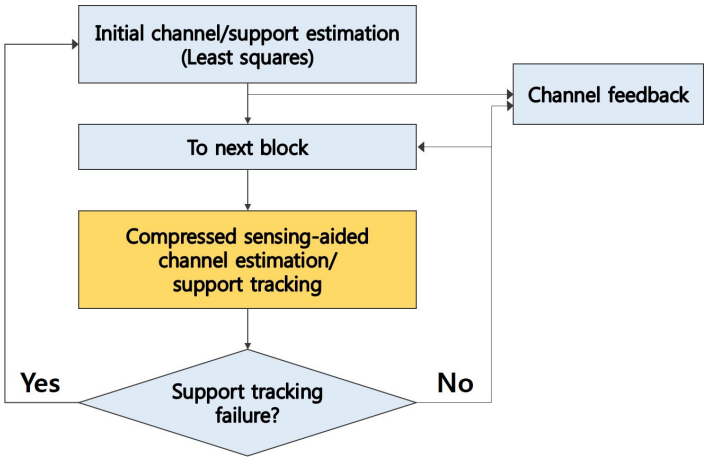

Figure 2.3: Pilot overhead reduction exploiting slowly varying support.

2.3 CS-Aided Channel Training

The slowly varying support model, described in the previous section, can be exploited to reduce the training overhead. For the overhead reduction, I introduce a support tracking concept whose operation is described in Fig. 2.3. While at least M channel uses are required at the initial stage where no prior knowledge is available, the succeeding block channel can be trained with a reduced overhead by utilizing the support extracted from the previous estimation. Note that a length M training sequence can be used initially or when a complete tracking failure of the support occurs.

The key component for the proposed framework is the estimation scheme exploiting the previous support information. To reduce the training overhead, I propose a CS-aided channel training scheme that utilizes the LS and CS approaches simultaneously. Based on the observation that the support of the transformed channel vector changes very slowly, the proposed scheme separates the channel into two vectors based on the previous support and estimates them with different strategies. In the following, I assume that the channel sparsity parameters k and k_s are known. In practice, k and

k_s can be obtained by observing long-term estimations [38], or a learning-based algorithm can be utilized to adapt the values to the channel variation.

2.3.1 Training Sequence Design

To reduce the training overhead, the proposed scheme separates the angular domain channel based on the previous support and pursues different strategies for the estimation of separated channels. For the channel separation, the training sequence is divided into two parts. While a part of length k is dedicated for the sounding of coefficients of $\mathbf{s}_{i,d}$ defined in (2.10), the remaining part of length $T_p - k$ is utilized for the estimation of $\mathbf{s}_{i,s}$ defined in (2.11). [1] The pilot structure can be expressed as

$$\mathbf{X}_{i,T} = \begin{bmatrix} \mathbf{X}_{i,d} & \mathbf{X}_{i,s} \end{bmatrix}, \tag{2.16}$$

where $\mathbf{X}_{i,d} \in \mathbb{C}^{M \times k}$ is the training sequence for $\mathbf{s}_{i,d}$, and $\mathbf{X}_{i,s} \in \mathbb{C}^{M \times (T_p - k)}$ is for $\mathbf{s}_{i,s}$.

To estimate $\mathbf{s}_{i,d}$, the former part $\mathbf{X}_{i,d}$ is designed as

$$\mathbf{X}_{i,d} = \mathbf{\Psi}_{\Omega_{i-1}} \tilde{\mathbf{X}}_d, \tag{2.17}$$

where $\tilde{\mathbf{X}}_d \in \mathbb{C}^{k \times k}$ is an orthonormal matrix satisfying $\tilde{\mathbf{X}}_d^H \tilde{\mathbf{X}}_d = \mathbf{I}_k$. With this construction, the dense vector of length k can be estimated through the LS filtering since we have k measurements.

The latter part $\mathbf{X}_{i,s}$ is used for the estimation of $\mathbf{s}_{i,s}$. Since $\mathbf{s}_{i,s}$ is sparse according to the slowly varying support model, we can estimate it with a compressed training sequence whose length is $T_p - k < |\Omega_{i-1}^c| = M - k$. To estimate the sparse vector $\mathbf{s}_{i,s}$ with the CS approach, the training sequence is constructed as

$$\mathbf{X}_{i,s} = \mathbf{\Psi}_{\Omega_{i-1}^c} \mathbf{\Phi}^H, \tag{2.18}$$

[1] $T_p \geq k$ is assumed since channel estimation with a pilot of length $T_p < k$ is underdetermined even with the knowledge of the instantaneous support.

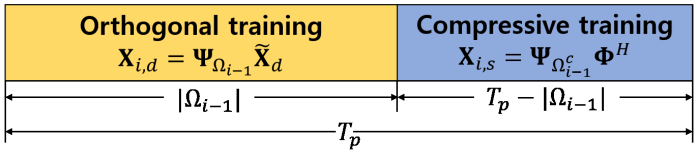

Figure 2.4: Pilot structure for the proposed channel estimation.

where $\mathbf{\Phi} \in \mathbb{C}^{(T_p-k)\times(M-k)}$ is a measurement matrix that satisfies the conditions for successful recovery.[2] The pilot structure for the channel separation is illustrated in Fig. 2.4.

2.3.2 Channel Estimation

With the proposed training sequence design, the received signal at the training phase is given by

$$\mathbf{y}_{i,T}^T = \begin{bmatrix} \mathbf{y}_{i,d}^T & \mathbf{y}_{i,s}^T \end{bmatrix}, \tag{2.19}$$

where

$$\begin{aligned} \mathbf{y}_{i,d}^T =& \sqrt{\rho_T} \mathbf{h}_i^H \mathbf{X}_{i,d} + \mathbf{z}_{i,d}^T \\ =& \sqrt{\rho_T} \mathbf{s}_i^H (\mathbf{I}_M)_{\Omega_{i-1}} \tilde{\mathbf{X}}_d + \mathbf{z}_{i,d}^T \\ =& \sqrt{\rho_T} \mathbf{s}_{i,d}^H \tilde{\mathbf{X}}_d + \mathbf{z}_{i,d}^T, \end{aligned} \tag{2.20}$$

and

$$\begin{aligned} \mathbf{y}_{i,s}^T =& \sqrt{\rho_T} \mathbf{h}_i^H \mathbf{X}_{i,s} + \mathbf{z}_{i,s}^T \\ =& \sqrt{\rho_T} \mathbf{s}_i^H (\mathbf{I}_M)_{\Omega_{i-1}^c} \mathbf{\Phi}^H + \mathbf{z}_{i,s}^T \\ =& \sqrt{\rho_T} \mathbf{s}_{i,s}^H \mathbf{\Phi}^H + \mathbf{z}_{i,s}^T. \end{aligned} \tag{2.21}$$

[2]The sufficient condition of measurement matrix which guarantees the support recovery can be characterized in terms of various criteria. Among them, the RIP condition is adopted in this chapter. Readers who are interested in other conditions are referred to [22–24] and references therein.

The vector $\mathbf{s}_{i,d}$ can be estimated from $\mathbf{y}_{i,d}$. With the LS approach, the estimate of $\mathbf{s}_{i,d}$ is obtained as

$$\tilde{\mathbf{s}}_{i,d} = \frac{1}{\sqrt{\rho_T}}\tilde{\mathbf{X}}_d \mathbf{y}_{i,d}^* = \mathbf{s}_{i,d} + \frac{1}{\sqrt{\rho_T}}\tilde{\mathbf{X}}_d \mathbf{z}_{i,d}^*, \tag{2.22}$$

where $\tilde{\mathbf{s}}_{i,d}$ denotes the estimate of $\mathbf{s}_{i,d}$. On the other hand, $\tilde{\mathbf{s}}_{i,s}$, the estimate of $\mathbf{s}_{i,s}$, can be obtained from $\mathbf{y}_{i,s}$ using a sparse recovery algorithm. Among various algorithms, I focus on the OMP algorithm [39] whose procedures are as follows:

1. Initialize the residual $\mathbf{r}_0 = \mathbf{y}_{i,s}^*$, the index set $\Lambda_0 = \emptyset$, and the iteration counter $t = 1$.

2. Find the index of the column of $\mathbf{\Phi}$ which has the maximum correlation with the residual $\mathbf{r}_{t-1}$, i.e., $j_t = \underset{j \in \{1,2,\ldots,M-k\}}{\operatorname{argmax}} |\boldsymbol{\phi}_j^H \mathbf{r}_{t-1}|$, where $\boldsymbol{\phi}_j$ denotes the jth column of the measurement matrix $\mathbf{\Phi}$.

3. Update the index set $\Lambda_t = \Lambda_{t-1} \cup \{j_t\}$.

4. Estimate the signal based on the current index set $\mathbf{u}_t = (\mathbf{I}_{M-k})_{\Lambda_t} \left(\mathbf{\Phi}_{\Lambda_t}^H \mathbf{\Phi}_{\Lambda_t}\right)^{-1} \mathbf{\Phi}_{\Lambda_t}^H \mathbf{y}_{i,s}^*$ and update the residual $\mathbf{r}_t = \mathbf{y}_{i,s}^* - \mathbf{\Phi}\mathbf{u}_t$.

5. If $t = k_s$, stop and estimate $\tilde{\mathbf{s}}_{i,s} = \frac{1}{\sqrt{\rho_T}}\mathbf{u}_t$. Otherwise, update $t = t + 1$ and return to Step 2.

Note that if the magnitude of the residual signal after k_s iterations is larger than a predefined threshold, we consider it as a support tracking failure and perform an initial channel estimation with a long sequence, as described in Fig. 2.3.

With the estimates $\tilde{\mathbf{s}}_{i,d}$ and $\tilde{\mathbf{s}}_{i,s}$, a primary estimate can be obtained as

$$\tilde{\mathbf{s}}_i = (\mathbf{I}_M)_{\Omega_{i-1}} \tilde{\mathbf{s}}_{i,d} + (\mathbf{I}_M)_{\Omega_{i-1}^c} \tilde{\mathbf{s}}_{i,s}. \tag{2.23}$$

Note that $\tilde{\mathbf{s}}_i$ can have up to $k + k_s$ nonzero elements due to the nature of the OMP process, while $\mathbf{s}_i$ has at most k nonzero elements. To refine the estimate and obtain the

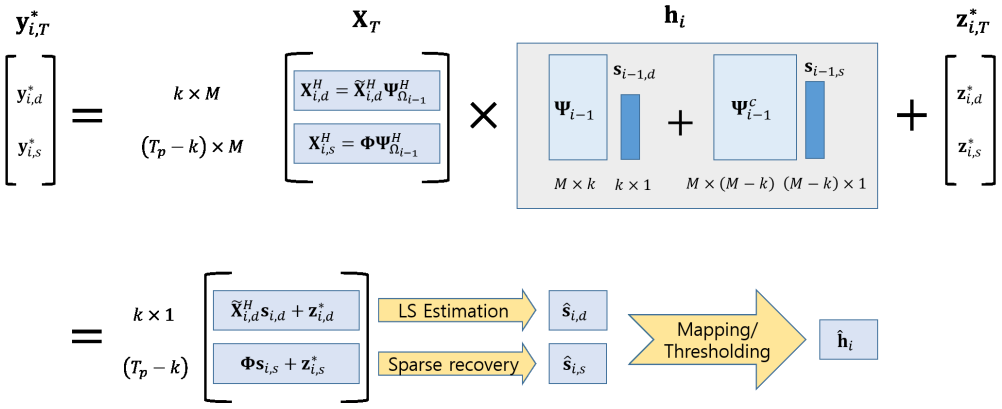

Figure 2.5: Illustration of the proposed channel estimation procedure.

support set for the estimation of the next block channel, we construct a final estimate of the channel by selecting indices corresponding to the k largest absolute values of $\hat{\mathbf{s}}_i$, i.e.,

$$\hat{\mathbf{s}}_i = \underset{\mathbf{s}\in\mathbb{C}^{M\times 1}, \|\mathbf{s}\|_0 \le k}{\operatorname{argmin}} \|\mathbf{s} - \tilde{\mathbf{s}}_i\|_1. \tag{2.24}$$

When OMP is used for support recovery in the presence of noise, the stopping rule in Step 5 generally includes the norm of the residual, i.e., stop if $t = k_s$ or $\|\mathbf{r}_t\|_2 < \eta$, where η is threshold determined by the noise level [40, 41]. In the proposed scheme, however, the iteration continues until $t = k_s$. Instead of thresholding, the support is pruned as (2.24) after combining the results of LS and CS estimation. The overall procedure of the proposed channel estimation scheme, along with the construction of pilot sequence, is illustrated in Fig. 2.5.

2.3.3 Estimation Error

In this subsection, the estimation error of the proposed scheme is analyzed in terms of the mean squared error (MSE), which is defined as

$$MSE = \mathbb{E}\left[\|\mathbf{e}_i\|_2^2\right] = \mathbb{E}\left[\|\mathbf{h}_i - \hat{\mathbf{h}}_i\|_2^2\right], \tag{2.25}$$

where $\mathbf{e}_i = \mathbf{h}_i - \hat{\mathbf{h}}_i$ denotes the estimation error vector. For analytical convenience, I assume that supports of consecutive channel blocks have exactly k_s different elements, i.e., $|\Omega_i \setminus \Omega_{i-1}| = |\Omega_{i-1} \setminus \Omega_i| = k_s$. Furthermore, I assume that the OMP algorithm is used for sparse recovery, and the measurement matrix $\mathbf{\Phi}$ is characterized in terms of the RIP, which is defined as below.

Definition 1 (Restricted Isometry Property [42])**.** A matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$ $(m < n)$ is said to satisfy the restricted isometry property (RIP) of order K with an isometry constant $\delta_K \in (0, 1)$ if

$$(1 - \delta_K)\|\mathbf{v}\|_2^2 \leq \|\mathbf{A}\mathbf{v}\|_2^2 \leq (1 + \delta_K)\|\mathbf{v}\|_2^2 \tag{2.26}$$

holds for all $\mathbf{v} \in \mathbb{C}^{n \times 1}$ such that $\|\mathbf{v}\|_0 \leq K$, where $\|\mathbf{v}\|_0$ denotes the number of nonzero elements of $\mathbf{v}$.

By leveraging the proposed estimation procedure and the sufficient condition for the support recovery, which has been developed in [41], I obtain the following results about the MSE.

Proposition 1. Assume that the SNR ρ_T is sufficiently large. When the OMP algorithm is utilized and the scaled measurement matrix $\bar{\mathbf{\Phi}} = \sqrt{\frac{M-k}{T_p-k}}\mathbf{\Phi}$ satisfies the RIP of order $k_s + 1$ with isometry constant $\delta_{k_s+1} < \frac{1}{\sqrt{k_s}+1}$, the induced MSE is given as

$$MSE = \frac{k - k_s}{\rho_T} + \frac{1}{\rho_T}\mathsf{tr}\left(\left(\mathbf{\Phi}_{\Lambda_{i,s}}^H \mathbf{\Phi}_{\Lambda_{i,s}}\right)^{-1}\right), \tag{2.27}$$

where $\Lambda_{i,s} \subset \{1, 2, \ldots, M - k\}$ is the support of $\mathbf{s}_{i,s}$. The MSE is bounded as

$$\begin{aligned} \frac{1}{\rho_T}\left(k - k_s + \frac{(M-k)k_s}{(T_p - k)(1 + \delta_{k_s+1})}\right) &\leq MSE \\ &\leq \frac{1}{\rho_T}\left(k - k_s + \frac{(M-k)k_s}{(T_p - k)(1 - \delta_{k_s+1})}\right). \end{aligned} \tag{2.28}$$

Proof. Following the definitions, the MSE given the support can be decomposed as

$$\begin{aligned}\text{MSE} =&\mathbb{E}\left[\|\mathbf{h}_i - \hat{\mathbf{h}}_i\|_2^2\right] \\ =&\mathbb{E}\left[\|\boldsymbol{\Psi}_{\Omega_{i-1}}(\mathbf{s}_{i,d} - \hat{\mathbf{s}}_{i,d}) + \boldsymbol{\Psi}_{\Omega_{i-1}^c}(\mathbf{s}_{i,s} - \hat{\mathbf{s}}_{i,s})\|_2^2\right] \\ =&\mathbb{E}\left[\|\mathbf{s}_{i,d} - \hat{\mathbf{s}}_{i,d}\|_2^2\right] + \mathbb{E}\left[\|\mathbf{s}_{i,s} - \hat{\mathbf{s}}_{i,s}\|_2^2\right].\end{aligned} \tag{2.29}$$

The first term can be expressed as

$$\begin{aligned}\mathbb{E}\left[\|\mathbf{s}_{i,d} - \hat{\mathbf{s}}_{i,d}\|_2^2\right] =&\mathbb{E}\left[\|(\mathbf{s}_{i,d} - \tilde{\mathbf{s}}_{i,d})_{\Lambda_{i,d}}\|_2^2\right] \\ =&\frac{1}{\rho_T}\mathbb{E}\left[\|(\tilde{\mathbf{X}}_d \mathbf{z}_{i,d}^*)_{\Lambda_{i,d}}\|_2^2\right] \\ =&\frac{k-k_s}{\rho_T},\end{aligned} \tag{2.30}$$

using (2.22).

For the second term, recall that $\hat{\mathbf{s}}_{i,s}$ is the output of the OMP algorithm. According to the result of [41], the support of $\mathbf{s}_{i,d}$ is exactly recovered by OMP under the assumption of RIP and sufficiently large SNR. Given the exact support recovery, the OMP reconstructs the sparse signal with the LS approach. As a result, the output of the OMP becomes

$$\hat{\mathbf{s}}_{i,s} = \frac{1}{\sqrt{\rho_T}}(\mathbf{I}_{M-k})_{\Lambda_{i,s}}\left(\boldsymbol{\Phi}_{\Lambda_{i,s}}^H \boldsymbol{\Phi}_{\Lambda_{i,s}}\right)^{-1}\boldsymbol{\Phi}_{\Lambda_{i,s}}^H \mathbf{y}_{i,s}^*. \tag{2.31}$$

With this, we can write the second term in (2.29) as

$$\begin{aligned}&\mathbb{E}\left[\|\mathbf{s}_{i,s} - \hat{\mathbf{s}}_{i,s}\|_2^2\right] \\ &= \frac{1}{\rho_T}\mathbb{E}\left[\|(\mathbf{I}_{M-k})_{\Lambda_{i,s}}(\boldsymbol{\Phi}_{\Lambda_{i,s}}^H \boldsymbol{\Phi}_{\Lambda_{i,s}})^{-1}\boldsymbol{\Phi}_{\Lambda_i}^H \mathbf{z}_{i,s}^*\|_2^2\right] \\ &= \frac{1}{\rho_T}\mathsf{tr}\big((\boldsymbol{\Phi}_{\Lambda_{i,s}}^H \boldsymbol{\Phi}_{\Lambda_{i,s}})^{-1}\big).\end{aligned} \tag{2.32}$$

Combining (2.30) and (2.32), the equality of (2.27) can be obtained.

With a trace property, we have

$$\mathsf{tr}\big((\boldsymbol{\Phi}_{\Lambda_i}^H \boldsymbol{\Phi}_{\Lambda_i})^{-1}\big) = \sum_{j=1}^{k_s}\frac{1}{\lambda_j}, \tag{2.33}$$

where the λ_j's are the eigenvalues of $\mathbf{\Phi}_{\Lambda_i}^H \mathbf{\Phi}_{\Lambda_i}$. Since $\bar{\mathbf{\Phi}}$ satisfies the RIP of order k_s+1, it is obvious that the eigenvalues are bounded as

$$\frac{T_p - k}{M - k}(1 - \delta_{k_s+1}) \leq \lambda_j \leq \frac{T_p - k}{M - k}(1 + \delta_{k_s+1}), \quad \text{for all } j, \tag{2.34}$$

resulting in

$$\begin{aligned} \frac{(M-k)k_s}{\rho_T(T_p - k)(1 + \delta_{k_s+1})} &\leq \mathbb{E}\left[\|\mathbf{s}_{i,s} - \hat{\mathbf{s}}_{i,s}\|_2^2\right] \\ &\leq \frac{(M-k)k_s}{\rho_T(T_p - k)(1 - \delta_{k_s+1})}. \end{aligned} \tag{2.35}$$

Using (2.30) and (2.35), we can obtain the bound of MSE. □

From the result, we can see that the upper bound in (2.28) increases as k_s increases for a fixed $T_p \leq M$ since the isometry constant of a matrix increases with its order. This implies that as the consecutive channels share larger part of support (i.e., smaller k_s), the proposed scheme can estimate the channel with smaller error, validating the efficient use of the prior support information.

2.4 Discussions

2.4.1 Design of Measurement Matrix

As seen in the previous section, the estimation performance of the proposed scheme highly depends on the design of the measurement matrix $\mathbf{\Phi}$. Specifically, the support can be recovered with higher probability and estimation error becomes smaller as the isometry constant decreases. For these reasons, the design of a matrix satisfying the RIP with a small isometry constant has been of great interest in various CS applications [21]. Since finding a matrix with the smallest isometry constant is infeasible due to the combinatorial nature of the RIP [43], most CS research has considered randomly generated sensing matrices. For example, it has been shown that a matrix with

i.i.d. random entries drawn from a Gaussian distribution [44] or a matrix designed by collecting random rows of a unitary matrix (e.g., a DFT matrix) can satisfy the RIP with a high probability. Specifically, it is known that a Gaussian matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$ satisfies RIP of order K with an overwhelming probability if $m \geq C \cdot \log(n/K)$, where C is a constant depending on the isometry constant δ_K [21]. Based on the result, we can construct a pilot signal according to the process of (2.16)-(2.18) with a randomly generated measurement matrix $\mathbf{\Phi}$. On the other hand, an optimization approach for measurement matrix design is also being actively studied by establishing objective functions and developing corresponding algorithms [45, 46]. Although there is no performance measure that is directly related to the CS performance, adopting a measurement matrix developed by these algorithms as $\mathbf{\Phi}$ can improve the estimation performance.

2.4.2 Extension to MIMO Systems

Throughout the chapter, I have considered MISO systems where the receiver is equipped with a single antenna. When multiple antennas are deployed at the receiver, the input-output relation of (2.1) is replaced by

$$\mathbf{y}[n] = \sqrt{\rho}\mathbf{H}_i^H \mathbf{x}[n] + \mathbf{z}[n], \tag{2.36}$$

where $\mathbf{H}_i \in \mathbb{C}^{M \times N}$ denotes the channel matrix between the transmitter and the receiver, N is the number of receive antennas, and $\mathbf{z}[n] \in \mathbb{C}^{N \times 1}$ is the Gaussian noise vector. In the MIMO setting, the channel estimation is the problem of estimating $\mathbf{H}$ with the received training signal

$$\mathbf{Y}_{i,T}^T = \sqrt{\rho_T}\mathbf{H}_i^H \mathbf{X}_T + \mathbf{Z}_{i,T}^T. \tag{2.37}$$

While the estimation of $\mathbf{H}_i$ can be performed column-wise operation, the estimation performance can be improved by exploiting the joint sparsity of MIMO channels.

With a transformation similar to the MISO case, the channel matrix can be represented in the angular domain [30], i.e.,

$$\mathbf{H}_i = \mathbf{\Psi}_t \mathbf{S}_i \mathbf{\Psi}_r^H, \tag{2.38}$$

where $\mathbf{\Psi}_t \in \mathbb{C}^{M\times M}$ and $\mathbf{\Psi}_r \in \mathbb{C}^{N\times N}$ denote the angular transform matrices at the transmitter and the receiver, respectively. Due to the relatively rich scattering experienced by the receiver, the columns of $\mathbf{S}_i$ usually have the same support [30, 47]. Hence, the proposed pilot design method, which exploits the support information of the previous block, is effective for systems with multiple receive antennas. Moreover, the joint sparsity of the columns can be utilized to enhance the estimation accuracy. The recovery of multiple sparse vectors with an identical support has been extensively studied [48, 49], and it was shown that the recovery performance can be improved by exploiting the joint sparsity. Consequently, the performance of the proposed scheme can be further improved in MIMO systems by utilizing a recovery algorithm developed for simultaneous sparse signal, which will be seen in Sec 2.5.

2.4.3 Comparison to CS with Partial Support Information

Since a slowly-varying support can be found in various applications where sequential sparse signals are reconstructed, attempts to exploit the partially known support have naturally arisen. Particularly, several CS algorithms that utilize partially known support information in sparse recovery have recently been introduced [38, 50–53]. In [50], a simple approach that applying CS on the LS residual computed using the previous support was proposed as a first CS solution utilizing partial support information. In [51], another solution, referred to as *modified-CS*, was proposed by the same authors, which finds a support that contains the smallest number of additional elements to the prior information. The modified-CS was further developed into *weighted* ℓ_1 *minimization* in [52] by taking into account the expected values of nonzero elements.

In [38], a greedy pursuit-based approach, referred to as *modified subspace pursuit (M-SP)*, was proposed to incorporate the partial support information. An approximate message passing algorithm [54] was also modified to exploit the partially known support [53]. Moreover, M-SP [38] and weighted ℓ_1 minimization [55] have been applied to the estimation of massive MIMO channels with a motivation similar to my work.

Hence, it is worth comparing the proposed channel estimation scheme with existing CS algorithms which incorporate prior support information. The key difference is that the proposed scheme utilizes prior support information to construct the pilot signal, while existing algorithms utilize the information only during their recovery process. While various CS algorithms have been modified to incorporate the partial support information in the recovery process, none of the previous work considers the adaptation of the measurement matrix to the best of my knowledge. Since the proposed scheme adapts the pilot sequence, it can estimate channels with more relaxed support condition and shows better estimation performance. The superior performance of the proposed scheme is numerically verified in Sec. 2.5. Moreover, the proposed scheme can be combined with various sparse recovery algorithms such as convex optimization or message passing algorithms to enhance the performance. However, due to the difference, the application of the proposed scheme is restricted to systems with a single user or multiple users with a common support while the existing recovery algorithms can be applied to a multi-user massive MIMO channel estimation.

2.5 Simulation Results

In this section, the performance of the proposed scheme is compared with several estimation schemes. For the proposed scheme, the measurement matrix $\mathbf{\Phi}$ is randomly drawn from the i.i.d. Gaussian distribution and normalized to satisfy the power constraint $\mathsf{tr}\left(\mathbf{\Phi}\mathbf{\Phi}^H\right) = T_p - k$. As a sparse recovery algorithms, OMP is used. The base-

line schemes are as follows.

- **Genie-aided LS:** The support Ω_i is assumed to be known. Using the information, the pilot is designed to train the subspace the support spans, and LS estimation is used. This ideal scheme provides a lower bound for MSE.

- **Static LS:** Only the support of the first block is known. The pilot is designed using the outdated support Ω_1 for all time and LS is used for estimation.

- **Random LS:** A randomly generated orthonormal sequence of length T_p is used for training and LS is used as an estimation filter.

- **OMP:** A random Gaussian matrix of size $M \times T_p$ is used for training signal. The OMP algorithm recovers the channel.

- **M-SP:** A random Gaussian matrix of size $M \times T_p$ is used for training signal. The M-SP algorithm, introduced in [38], recovers the channel incorporating the previous support information. Similarly to the proposed scheme, M-SP successively updates the support information.

First, I consider a toy channel model. In this model, an angular channel vector is assumed to have exactly k nonzero elements, while the remaining $M-k$ elements are zero. The support is randomly selected such that $k-k_s$ indices are preserved from the previous support. The nonzero coefficients are drawn from an i.i.d. complex Gaussian distribution with zero mean and unit variance. For the simulation, I set $M = 100$, $k = 40$, and $k_s = 3$.

Fig. 2.6 and Fig. 2.7 show the normalized mean squared error (NMSE) $\frac{\mathbb{E}\|\mathbf{h}_i - \hat{\mathbf{h}}_i\|_2^2]}{\mathbb{E}[\|\mathbf{h}_i\|_2^2}$ and the normalized beamforming gain $\mathbb{E}\left[\frac{|\mathbf{h}_i^H \hat{\mathbf{h}}_i|^2}{\|\mathbf{h}_i\|_2 \|\hat{\mathbf{h}}_i\|_2}\right]$ of various schemes, respectively, where the training is performed with the SNR of $\rho_T = 20$ [dB], and the length of training sequence varies from k to M. As expected, the low degrees of freedom of

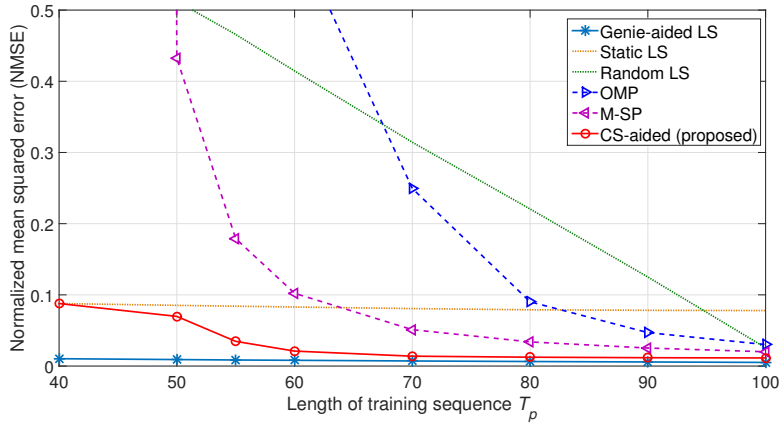

Figure 2.6: NMSE versus the length of the training sequence with $M = 100$, $k = 40$, $k_s = 3$, and SNR$= 20$ dB.

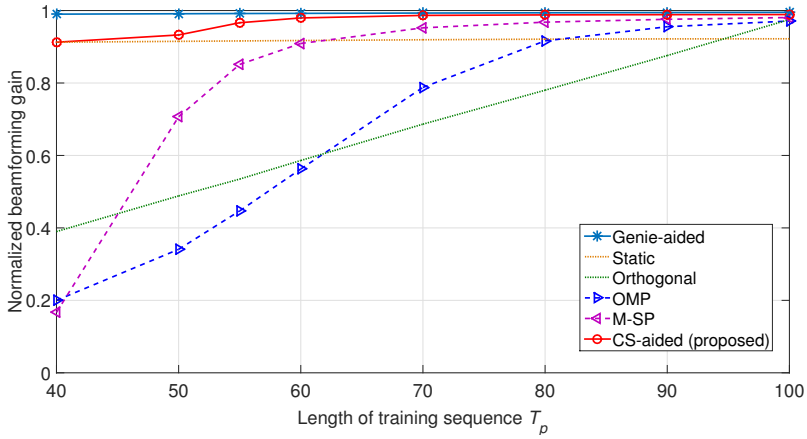

Figure 2.7: Normalized beamforming gain versus the length of the training sequence with $M = 100$, $k = 40$, $k_s = 3$, and (training) SNR$= 20$ dB.

the channel cannot be exploited using the random orthogonal pilot. OMP also fails to estimate with a reduced pilot overhead. This implies that the sparsity level of the channel (40 of 100 elements are nonzero) is not enough to be recovered by the CS approach. Since the support changes slowly, a static approach shows a smaller error

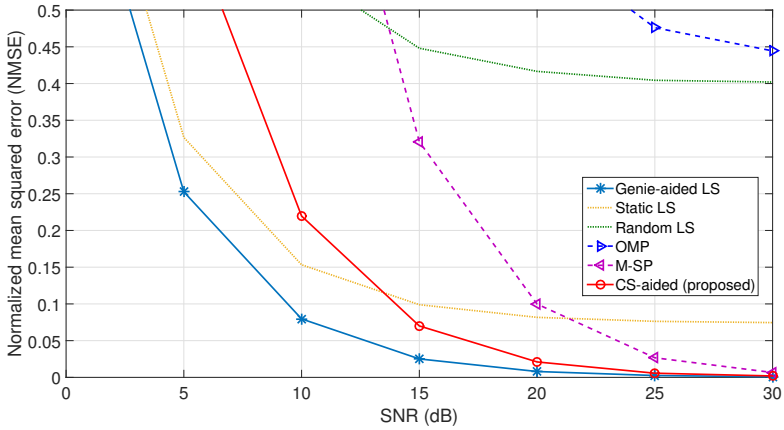

Figure 2.8: NMSE versus SNR with $M = 100$, $k = 40$, $k_s = 3$, and $T_p = 60$.

compared to the random approaches such as orthogonal sounding and OMP. However, the coefficients corresponding to the changed k_s support elements cannot be estimated with the static approach. M-SP, which exploits the previous support information in the recovery process, shows better performance than the conventional OMP. The proposed CS-aided channel estimation, which can estimate the sparse vector corresponding to the complement of the previous support by adding a short training sequence (about 20 additional symbols), outperforms the M-SP and shows a performance close to the ideal Genie-aided scheme. This results show the effectiveness of the channel separation and partial application of a CS technique.

Fig. 2.8 shows the NMSE of the estimation schemes when the training is performed with varying SNR. The length of the training sequence is set to $T_p = 60$. Since the CS algorithms is inherently sensitive to noise level, the estimation schemes that utilize CS, including the proposed scheme, show poor performance in low SNR region. On the other hand, the LS based schemes (Genie-aided and Static) show relatively low error. With an SNR larger than 15 dB, the proposed scheme outperforms the Static LS and shows reasonable performance. In all region, the proposed scheme shows lower

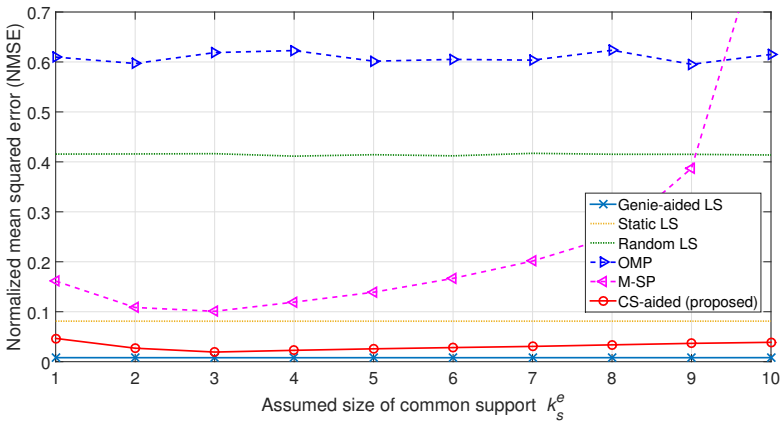

Figure 2.9: NMSE versus the mismatched parameter k_s^e with $M = 100$, $k = 40$, $k_s = 3$, and SNR$=$ 20 dB.

error than the conventional CS algorithms.

Thus far, I assumed that the channel evolution parameter k_s is exactly known. In practical implementations, the assumption may be unrealistic. Fig. 2.9 shows the NMSE performance when wrong value of k_s^e is known. I consider mismatched parameters $1 \leq k_s^e \leq 10$, while the real channel evolves with $k_s = 3$. As expected, estimation schemes that exploit the previous support information (M-SP and CS-aided) show performance degradation when the parameter mismatch occurs. However, the degradation of the proposed scheme is marginal, while M-SP is highly vulnerable to the mismatch. The robustness of the proposed scheme can be explained with two reasons. First, since only a partial part of the channel is estimated with a CS algorithm, the impact of the mismatch is limited. Moreover, even though wrong indices are selected by the CS algorithm operating with $k_s^e > k_s$, the pruning step can eliminate the effect of wrong indices.

In the following, the estimation performance is evaluated in a realistic channel evolution scenario. A ray-based channel model for uniform array antenna configura-

tion, which is used in [56], is adopted. According to the model, a channel vector of $M_v \times M_h$ UPA is constructed as

$$\mathbf{h} = \sum_{k=1}^{L} \frac{e^{j\varphi_k}}{\sqrt{L}} \mathbf{a}_h(u_k) \otimes \mathbf{a}_v(v_k), \tag{2.39}$$

where L is the number of independent paths, φ_k is a random phase for the kth path, $\mathbf{a}_h(u_k)$ and $\mathbf{a}_v(v_k)$ are antenna response vectors in horizontal and vertical axes, defined as

$$\mathbf{a}_h(u_k) = [1, e^{-ju_k}, \ldots, e^{-j(M_h-1)u_k}], \tag{2.40}$$

and

$$\mathbf{a}_v(v_k) = [1, e^{-jv_k}, \ldots, e^{-j(M_v-1)v_k}]. \tag{2.41}$$

u_k and v_k are defined as

$$u_k = \frac{2\pi D_h}{\lambda} \sin(\theta_k) \cos(\phi_k), \tag{2.42}$$

and

$$v_k = \frac{2\pi D_v}{\lambda} \cos(\theta_k), \tag{2.43}$$

where D_h and D_h are antenna spacing in horizontal/vertical axes, λ is the carrier wavelength, and θ_k and ϕ_k are vertical and horizontal angles of departure (AoD). The AoD pair of each path is assume to be normal distributed as $\theta_k \sim \mathcal{N}(\theta, \xi)$ and $\phi_k \sim \mathcal{N}(\phi, \sigma)$ with center angles θ, ϕ and angular spreads ξ, σ. In the simulation, I assumed $L = 15$ paths. To model the slow variation in scattering environment, I assume that one path disappears and another randomly generated path appears every fading block, while the total number of paths remains unchanged. Note that random phase φ_k is independently generated every block in this model.

Fig. 2.10 and Fig. 2.11 show the NMSE of various schemes when the channel evolves according to this model. In Fig. 2.10, a horizontal ULA of $M = 128$ antennas

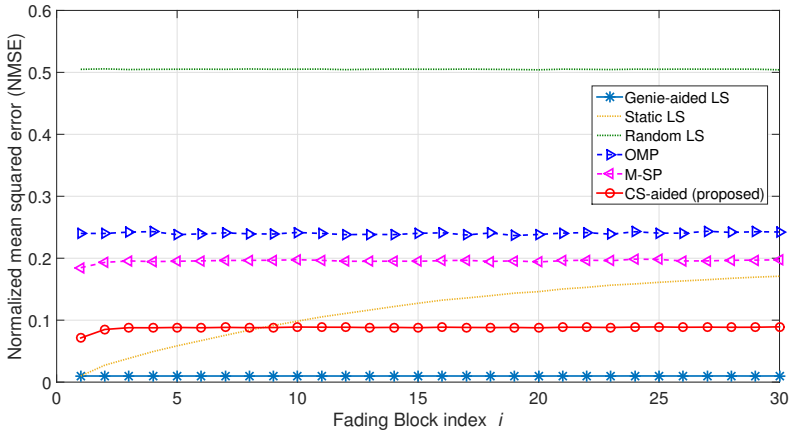

Figure 2.10: NMSE behavior according to the scattering evolution with a ULA of $M = 128$, $T_p = 64$, and SNR$= 20$ dB.

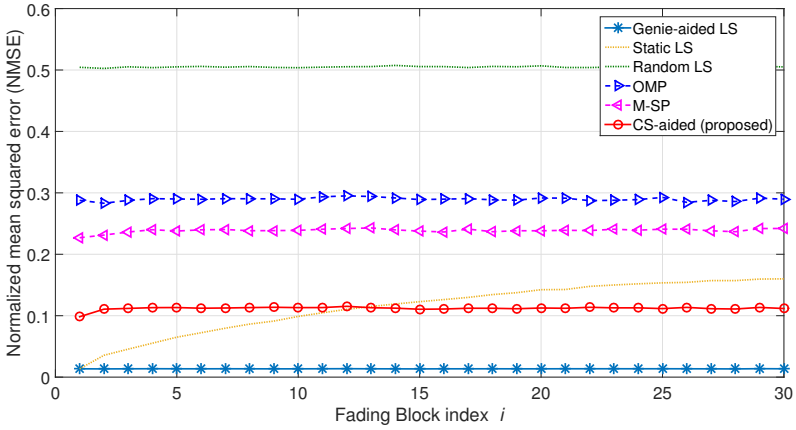

Figure 2.11: NMSE behavior according to the scattering evolution with a UPA of $(M_h \times M_v) = (32 \times 8)$, $T_p = 128$, and SNR$= 20$ dB.

is considered with antenna spacing $D_h = \lambda/2$. The mean and standard deviation of azimuth AoD are set to $\phi = \pi/3$ and $\sigma = \pi/6$, respectively. The DFT matrix is used for the angular domain transformation. For each block, $T_p = M/2 = 64$ channel uses are allocated for training signal, and the parameters for the proposed scheme are set as

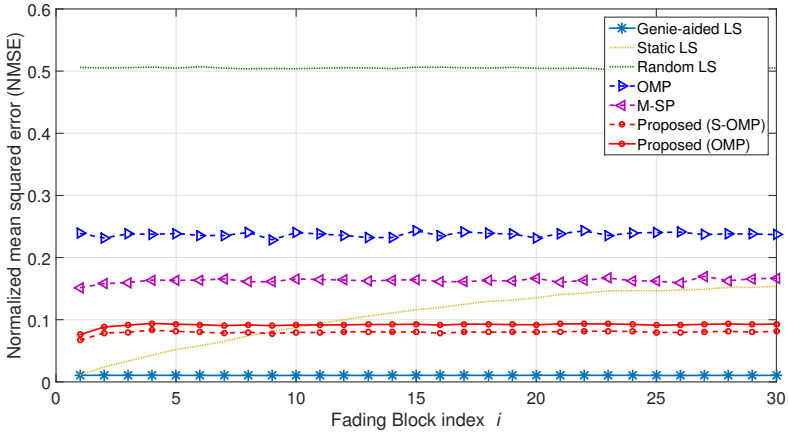

Figure 2.12: NMSE behavior according to the scattering evolution with $M = 128$ transmit antennas in ULA, $N = 2$ receive anennas, $T_p = 64$, and SNR$= 20$ dB.

$k = 32$ and $k_s = 10$. In Fig. 2.11, a UPA of $(M_h \times M_v) = (32 \times 8)$ is considered with antenna spacing of $D_h = \lambda/2$ and $D_v = \lambda$. The mean and standard deviation of elevation AoD are set to $\theta = \pi/4$ and $\xi = \pi/12$, respectively, while azimuth AoD of each path follows the same distribution as the ULA setting. The 2D-DFT is used for the angular domain transformation. The length of training sequence and parameters are set as $T_p = 128$, $k = 64$, and $k_s = 15$. For both scenarios, the SNR is set to $\rho_T = 20$ [dB]. The results show that both the random LS and the OMP fail to adequately estimate the channel. While M-SP, which utilizes the prior support information in its recovery process, shows a better performance compared to the conventional OMP, it still shows high estimation error. The error of the static approach utilizing the outdated support Ω_1 is kept small with a negligible change of the support, but increases as the scattering environment keeps changing. This implies that periodic update of channel statistics is required for the reduction of pilot, which can be a huge burden in massive MIMO systems. However, the proposed scheme shows stable error performance since it tracks the statistics without explicit update process.

Fig. 2.12 shows the estimation performance in a MIMO system with $N = 2$ receive antennas, in which the transmit side setting is the same as the ULA of Fig. 2.10. The angle-of-arrival (AoA) of each path at the receiver is assumed to be uniformly distributed in $[0, 2\pi]$. While most baseline schemes show performance similar to that of the MISO scenario, the performance of M-SP is improved since the algorithm implicitly utilizes the joint sparsity. The proposed scheme was performed with two sparse recovery algorithms. Solid line shows the performance with the OMP algorithm applied to each row, and dashed line denotes the performance with simultaneous OMP (S-OMP) exploiting the joint sparsity [49]. As predicted in 2.4.2, the performance is slightly improved by exploiting the joint sparsity. Since the proposed scheme can incorporate various sparse recovery algorithms, it is expected that the performance can be further improved by adopting a more advanced algorithm.

2.6 Conclusion

In this chapter, a CS-aided channel estimation scheme, which utilizes LS and CS simultaneously, was proposed for FDD massive MIMO systems. Under the assumptions that massive MIMO channels have fewer degrees of freedom than the number of transmit antennas, and the support of the angular domain channel changes slowly, the proposed scheme separates the channel into a dense vector and a sparse vector. By applying CS for the estimation of the sparse vector, the pilot overhead can be reduced when the channel is not sparse enough and conventional CS algorithms are not applicable. Numerical results verified that the proposed scheme can shorten training sequence significantly without explicit update of channel statistics.

Chapter 3

Projection-Based Differential Feedback

Channel state information at transmitter (CSIT) plays a key role in achieving potential gain of massive multiple-input multiple-output (MIMO) systems. In frequency division duplexing (FDD) systems, CSIT can be obtained through feedback from the receiver. Conventional limited feedback schemes, which have been designed for small-scale MIMO systems, suffer from prohibitive amount of feedback requirement and encoding complexity when the number of transmit antennas becomes massive. In this chapter, a projection-based differential feedback (PBDF) protocol is proposed for FDD massive MIMO systems. In the PBDF framework, a difference between original and predicted vectors is projected, and quantization is performed in a smaller dimensional subspace. With an appropriate projection exploiting spatial and temporal correlation of massive MIMO channels, the feedback amount and encoding complexity can be significantly reduced. The simulation results show that the proposed scheme achieves a large portion of potential throughput gain of massive MIMO systems with small amount of channel feedback.

3.1 Introduction

Over the last decade, multiple-input multiple-output (MIMO) technology has contributed to improvement in spectral efficiency and link reliability of wireless networks. Recently, massive MIMO systems, where base stations are equipped with a very large number of antennas, are drawing significant interests from both academia and industry. As the number of transmit antennas becomes larger, throughput and energy efficiency of the network can be unprecedentedly improved by constructing more precise spatial beams [4–6, 8]. Furthermore, such improvements can be accomplished by simple signal processing such as maximal ratio transmission (MRT) and maximal ratio combining (MRC). For these reasons, massive MIMO is considered as one of key technologies for next generation cellular networks [1].

To reap the aforementioned *massive MIMO gain*, however, accurate channel state information (CSI) at transmitter (CSIT) is essentially required. Due to the large dimension of a massive MIMO channel, most of previous studies have considered time division duplexing (TDD) systems where CSIT is achieved through uplink training by exploiting channel reciprocity [4–6]. In practice, however, sophisticated calibration process is required to achieve the channel reciprocity. Moreover, frequency division duplexing (FDD) is typically used in contemporary cellular networks. Accordingly, FDD is expected to be adopted at least for the early stage of massive MIMO systems, and CSIT acquisition for the FDD systems is of great interest.

In FDD systems, where the uplink-downlink channel reciprocity does not hold, the transmitter acquires the CSI via feedback from the receiver that estimates the CSI during training period. Due to the large dimensional channel, both of channel training and feedback are challenging in massive MIMO systems. Since conventional training schemes consume the downlink resource proportional to the number of antennas, several schemes have been recently developed to perform channel estimation with re-

duced training period [19, 20]. On the other hand, since the feedback of a large dimensional channel would consume a large portion of uplink resource, an efficient feedback scheme for massive MIMO systems is also needed.

Since accuracy of CSIT has a critical impact on the performance of multi-user MIMO systems, extensive studies have been devoted to limited feedback in conventional (small-scale) MIMO systems [57–60, 62–66]. For spatially uncorrelated channels, Grassmannian codebook was designed to maximize the minimum chordal distance between codewords [58], and performance of random vector quantization was analyzed in single-user [59] and multi-user MIMO systems [60]. In [62] and [63], codebooks for spatially correlated channels were developed by rotating an independent and identically distributed (i.i.d.) codebook to align with the channel distribution. For temporally correlated channels, differential feedback concept has been investigated, where a codebook is generated around the previous quantized vector [64–66]. To achieve the rate of the perfect CSIT within a constant gap with such schemes, the codebook size (in bits) needs to scale linearly with the number of the transmit antennas [60]. Hence, a large portion of the uplink resource is inevitably consumed for the CSI feedback to achieve the high spectral efficiency of massive MIMO systems. Moreover, encoding complexity increasing with the codebook size becomes a huge burden. Hence, a novel CSI feedback strategy that provides accurate CSIT with a low overhead and low encoding complexity is demanded for FDD massive MIMO systems.

For massive MIMO systems, a trellis code-based quantization scheme was developed capturing the duality between source encoding (quantization) and channel decoding in [15] and extended to spatially and temporally correlated channel in [67]. Although this approach reduces encoding complexity significantly, the required feedback amount is still prohibitive (more bits than the number of antennas) with a massive number of antennas. Another attractive way is to reduce the dimension of feed-

back [18,31,69–71], which is based on the observation that spatially correlated massive MIMO channels have degrees of freedom much smaller than the dimension of channels. In [68], a dimension reduction strategy based on grouping of highly correlated antenna elements was proposed. In [31] and [69], a compressive sensing scheme with random projection was applied to feedback. However, compressed sensing can be applied only for extremely correlated channels since the approach requires a sparse channel assumption with few nonzero elements in the angular domain. In [18] and [70], a new framework for FDD massive MIMO, referred to as joint spatial division and multiplexing (JSDM), was proposed under the assumption of clustered users. In JSDM, users are divided into groups such that the users of each group have similar covariance matrix, and channels are projected into the null-space of other group channel spaces to reduce the dimension of the feedback information. However, since covariance matrices of users within a group are required to be almost identical, JSDM can be effective only with a scheduling that selects users to be served from numerous candidates [70]. A joint design of training pilot and feedback strategy has been proposed for multi-user systems exploiting the spatial channel correlation in [71]. However, such dimension reduction schemes only exploit spatial correlation without considering temporal correlation.

In this chapter, a limited feedback strategy, referred to as projection-based differential feedback (PBDF), is proposed to achieve accurate CSIT with a low feedback overhead and low encoding complexity. To reduce the amount of feedback efficiently, I exploit both of spatial and temporal correlation.[1] Main contribution of this chapter is twofold as follows.

- I propose a PBDF framework for FDD massive MIMO systems. In the PBDF

[1] I assume that the correlation information, which is slowly varying, is known at both the transmitter and the receiver. This assumption can be realized using the methods explained in Sec. 3.4.2.

framework, a difference vector is projected into a smaller dimensional subspace, and the vector with reduced dimension is quantized and fed back to the transmitter. By combining differential feedback and projection concepts, accurate CSIT can be achieved with a small amount of feedback, and the complexity for codebook search is also significantly reduced. Furthermore, lower dimensional codebooks adopted in contemporary wireless standards (e.g. LTE, LTE-A, etc.) can be reused for massive MIMO systems enhancing the backward compatibility.

- The optimal projection matrix design for the PBDF framework is developed. The optimal projection matrix can be obtained by a novel update of difference vector distribution, followed by Karhunen-Loève transform (KLT). Exploiting channel statistics (spatial and temporal correlation) and quantizer characteristics jointly, the projection matrix can be calculated at the transmitter and the receiver simultaneously. The proposed update process can be performed with a low-complexity algorithm exploiting the eigenstructure of the covariance matrix. Furthermore, the update process is extended to imperfect CSI at the receiver (CSIR) scenario, considering that downlink channel training is also a challenging issue in massive MIMO systems [20], [19] .

The remainder of the chapter is organized as follows. In Section 3.2, the system model and massive MIMO channel model are introduced. In Section 3.3, the PBDF framework is proposed for FDD massive MIMO systems, along with a design procedure for the optimal projection matrix. Some practical issues for the implementation are discussed in Section 3.4, followed by numerical results in Section 3.5. Finally, conclusion is drawn in Section 3.6.

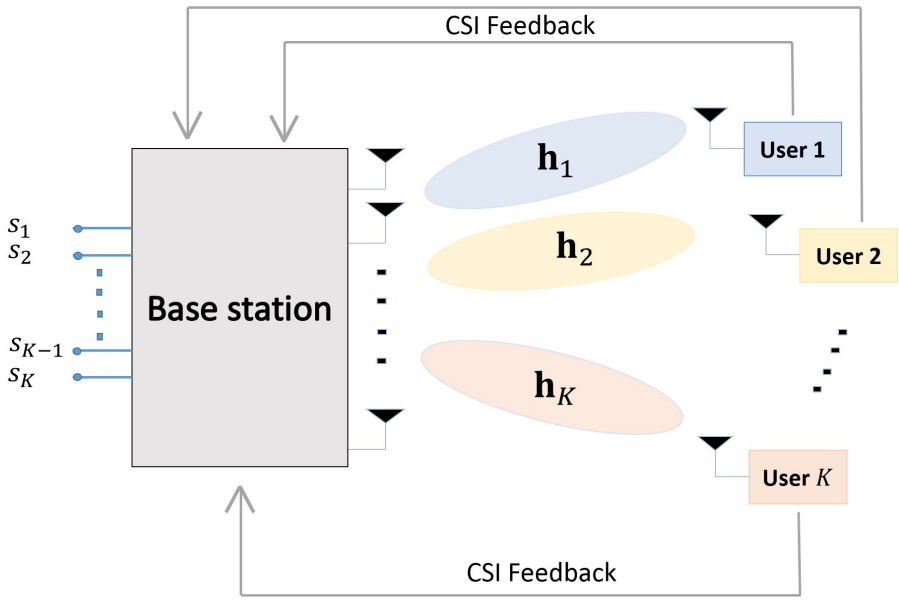

Figure 3.1: FDD multi-user massive MIMO system.

3.2 System Model

In this chapter, a single-cell multi-user massive MIMO system is considered. The system consists of a base station (BS) equipped with M antennas and K single-antenna users. The number of BS antennas M is assumed to be much greater than the number of users K ($M \gg K$), and no scheduling is considered. The received signal of user k at the nth symbol time is given by

$$y_k[n] = \mathbf{h}_k^H[n]\mathbf{x}[n] + z_k[n], \tag{3.1}$$

where $\mathbf{h}_k[n] \in \mathbb{C}^{M\times 1}$ is the channel vector at the nth symbol time, $\mathbf{x}[n] \in \mathbb{C}^{M\times 1}$ is the transmit signal vector at the nth symbol time with power constraint $\mathbb{E}[\|\mathbf{x}[n]\|_2^2] = P$, and $z_k[n] \in \mathbb{C}^{1\times 1}$ is additive noise of user k at the nth symbol time distributed according to $\mathcal{CN}(0, 1)$. The overall system is depicted in Fig. 3.1.

I assume the block fading, i.e., channel vectors remain constant for L consecutive symbol times. The channel vector of the kth user at the ith fading block is defined as

$$\mathbf{h}_{i,k} = \mathbf{h}_k \left[iL + l \right], \quad i \geq 1, \quad l \geq 0. \tag{3.2}$$

I assume that each user quantizes and sends the corresponding channel vector to the BS through a feedback link every L channel uses.

3.2.1 Multi-User Beamforming with Limited Feedback

When linear precoding is used with equal power allocation, the transmit signal vector $\mathbf{x}[n]$ is constructed as

$$\mathbf{x}[n] = \sqrt{\frac{P}{K}} \mathbf{W}_i \mathbf{s}[n], \tag{3.3}$$

where i is the index of the fading block including the nth symbol time ($(i-1)L \leq n < iL$); $\mathbf{s}[n] = [s_1[n], s_2[n], \cdots, s_K[n]]^T \in \mathbb{C}^{K \times 1}$ denotes the message signal vector at the nth symbol time; $\mathbf{W}_i = [\mathbf{w}_{i,1}, \mathbf{w}_{i,2}, \cdots, \mathbf{w}_{i,K}] \in \mathbb{C}^{M \times K}$ is the precoding matrix for the ith channel block, where $\|\mathbf{w}_{i,k}\|_2^2 = 1$ for all k. Note that the precoding matrix is updated for every channel block (not every symbol time) since it depends on a channel realization.

By applying the precoding, the received signal (3.1) can be expressed as

$$\begin{aligned} y_k[n] &= \sqrt{\frac{P}{K}} \mathbf{h}_{i,k}^H \mathbf{W}_i \mathbf{s}[n] + z_k[n] \\ &= \sqrt{\frac{P}{K}} \mathbf{h}_{i,k}^H \mathbf{w}_{i,k} s_k[n] \\ &\qquad + \sqrt{\frac{P}{K}} \mathbf{h}_{i,k}^H \sum_{j=1, j \neq k}^{K} \mathbf{w}_{i,j} s_j[n] + z_k[n], \end{aligned} \tag{3.4}$$

and the signal-to-interference-plus-noise ratio (SINR) at the ith channel block is resulted as

$$\text{SINR}_{i,k} = \frac{\frac{P}{K} |\mathbf{h}_{i,k}^H \mathbf{w}_{i,k}|^2}{\frac{P}{K} \sum_{j=1, j \neq k}^{K} |\mathbf{h}_{i,k}^H \mathbf{w}_{i,j}|^2 + 1}. \tag{3.5}$$

Corresponding achievable sum-rate of the ith channel block is given as

$$R_{i,sum} = \sum_{k=1}^{K} \log_2(1 + \text{SINR}_{i,k}). \tag{3.6}$$

To handle the inter-user interference and achieve a high sum-rate, the BS needs to perform precoding utilizing CSI. However, perfect CSIT is hardly achievable in practical FDD systems. Instead, the BS generates a precoding matrix using quantized CSI that is provided via feedback. To generate a precoding matrix, I adopt the well-known zero-forcing beamforming (ZFBF) strategy [60], [72], which generates beamforming vectors orthogonal to channels of unintended users. With ZFBF, precoding vectors are determined to be $\mathbf{w}_{i,k} = \frac{\bar{\mathbf{W}}_i^{\text{ZF}}(:,k)}{\|\bar{\mathbf{W}}_i^{\text{ZF}}(:,k)\|_2}$, where $\bar{\mathbf{W}}_i^{\text{ZF}} = \hat{\mathbf{H}}_i(\hat{\mathbf{H}}_i^H\hat{\mathbf{H}}_i)^{-1}$ is the pseudo-inverse of $\hat{\mathbf{H}}_i$, and $\hat{\mathbf{H}}_i = [\hat{\mathbf{h}}_{i,1}, \hat{\mathbf{h}}_{i,2}, \cdots, \hat{\mathbf{h}}_{i,K}]$ is an estimate of $\mathbf{H}_i = [\mathbf{h}_{i,1}, \mathbf{h}_{i,2}, \cdots, \mathbf{h}_{i,K}]$ at the BS.

In conventional limited feedback systems, the codeword most aligned to the channel vector $\mathbf{h}_{i,k}$ is selected from a predefined codebook $\mathcal{C}$ and directly used for $\hat{\mathbf{h}}_{i,k}$, i.e.,

$$\hat{\mathbf{h}}_{i,k} = \underset{\mathbf{c}\in\mathcal{C}}{\text{argmax}} \, |\mathbf{c}^H\mathbf{h}_{i,k}|. \tag{3.7}$$

Note that the quantization process requires $\mathcal{O}(M2^B)$ complex multiplications, where B is the size of codebook in bit, i.e., $|\mathcal{C}| = 2^B$. Furthermore, the number of bits B needs to grow linearly with the codebook dimension M to achieve the high rate of the massive MIMO [60]. Hence, the conventional codebook approach suffers from inaccuracy and huge search complexity in massive MIMO systems due to the large dimension of channel vectors. As a result, it is of great importance to reduce the amount of feedback while providing accurate CSI to the BS.

3.2.2 Massive MIMO Channel

To achieve an accurate CSIT with a reduced amount of feedback, I exploit the spatially/temporally correlated nature of massive MIMO channel. Due to a large number of antenna elements packed in a limited space, massive MIMO channels tend to be spatially correlated [4]. Furthermore, I assume a slow-varying channel such that the channel vectors of adjacent fading blocks are correlated.

A spatially and temporally correlated massive MIMO channel can be modeled by a spatial correlation matrix and a temporal correlation coefficient, which are assumed to be static in time. The spatial correlation matrix of user k is defined as

$$\mathbf{R}_k = \mathbb{E}[\mathbf{h}_{i,k}\mathbf{h}_{i,k}^H], \quad \text{for all } i, \tag{3.8}$$

and can be decomposed as

$$\mathbf{R}_k = \mathbf{U}_k\mathbf{\Lambda}_k\mathbf{U}_k^H, \tag{3.9}$$

where columns of an orthonormal matrix $\mathbf{U}_k \in \mathbb{C}^{M\times M}$ is the eigenvectors of $\mathbf{R}_k$, and $\mathbf{\Lambda}_k = \text{diag}([\lambda_{k,1}^2, \lambda_{k,2}^2, \cdots, \lambda_{k,M}^2])$ is the diagonal matrix whose elements are eigenvalues of $\mathbf{R}_k$. Without loss of generality, I assume $\mathsf{tr}(\mathbf{R}_k) = \sum_{m=1}^{M}\lambda_{k,m}^2 = M$ and $\lambda_{k,1} \geq \lambda_{k,2} \geq \cdots \geq \lambda_{k,M} \geq 0$ for all $1 \leq k \leq K$. The temporal correlation coefficient η_k ($0 \leq \eta_k \leq 1$) is defined as

$$\eta_k = \frac{\mathbb{E}[|\mathbf{h}_{i,k}^H\mathbf{h}_{i-1,k}|]}{\sqrt{\mathbb{E}[\|\mathbf{h}_{i,k}\|_2^2]\mathbb{E}[\|\mathbf{h}_{i-1,k}\|_2^2]}}. \tag{3.10}$$

With $\mathbf{R}_k$ and η_k, I assume that the channel follows Gauss-Markov model, i.e.,

$$\begin{aligned}\mathbf{h}_{1,k} =& \mathbf{U}_k\mathbf{\Lambda}_k^{\frac{1}{2}}\mathbf{g}_{1,k},\\ \mathbf{h}_{i,k} =& \eta_k\mathbf{h}_{i-1,k} + \sqrt{1-\eta_k^2}\mathbf{U}_k\mathbf{\Lambda}_k^{\frac{1}{2}}\mathbf{g}_{i,k}, \quad i \geq 2,\end{aligned} \tag{3.11}$$

where $\mathbf{g}_{i,k} \in \mathbb{C}^{M\times 1}$ is an innovation process for the ith block channel of user k, and $\mathbf{g}_{i,k} \sim \mathcal{CN}(0, \mathbf{I}_M)$.

As the existing works in the literature [18, 19], I assume that both the BS and the user know the channel statistics ($\mathbf{R}_k$ and η_k) throughout this chapter. In Section 3.4.2, methods realizing this assumption are discussed.

3.3 Projection-Based Differential Feedback

In this section, I present the PBDF framework, which exploits spatial and temporal correlation of massive MIMO channel to reduce the feedback overhead. Then, the design methodology that find the optimal projection matrix for the framework is also developed.

3.3.1 Projection-Based Differential Feedback Framework

The process of the BS and the user for CSIT acquisition in PBDF framework is described in Fig. 3.2. From here, the user index k is omitted for the sake of simplicity.[2] In this section, I assume perfect CSIR, i.e., the user knows the corresponding channel information exactly. The PBDF is extended to imperfect CSIR scenario in Section 3.4.2.

The PBDF procedure is as follows:

- The user obtains the corresponding channel information $\mathbf{h}_i$ through downlink training.
- (For $i \geq 2$,) The user generates a difference vector $\tilde{\mathbf{h}}_i$ by subtracting $\eta\hat{\mathbf{h}}_{i-1}$, which is the best prediction of $\mathbf{h}_i$ based on the previous reconstruction $\hat{\mathbf{h}}_{i-1}$ in

[2] In the PBDF framework, the compression and reconstruction of CSI are performed in a distributed manner, i.e., the procedure of one user is independent of the procedure of other users. Hence, I can drop the user index k for the simplicity.

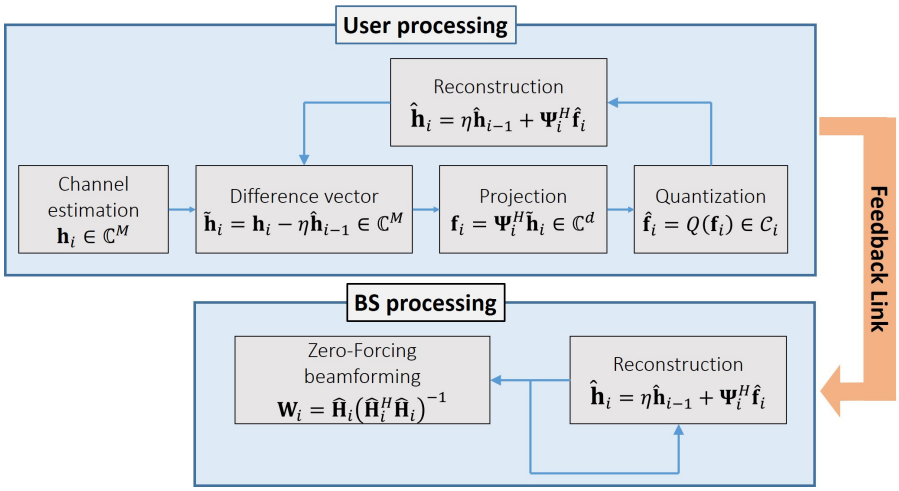

Figure 3.2: The block diagram of BS/user process for PBDF.

a mean squared error sense:

$$\tilde{\mathbf{h}}_i = \mathbf{h}_i - \eta\hat{\mathbf{h}}_{i-1}. \tag{3.12}$$

- The difference vector is projected into a lower dimensional subspace:

$$\mathbf{f}_i = \boldsymbol{\Psi}_i^H\tilde{\mathbf{h}}_i = \boldsymbol{\Psi}_i^H(\mathbf{h}_i - \eta\hat{\mathbf{h}}_{i-1}) \in \mathbb{C}^{d\times 1}, \tag{3.13}$$

where $\boldsymbol{\Psi}_i \in \mathbb{C}^{M\times d}$ is an orthonormal projection matrix (i.e. $\boldsymbol{\Psi}_i^H\boldsymbol{\Psi}_i = \mathbf{I}_d$), and d is the dimension of projection space. Note that d is a design parameter whose effect on feedback accuracy depends on the channel statistics. At $i = 1$, channel vector $\mathbf{h}_1$ itself is projected: $\mathbf{f}_1 = \boldsymbol{\Psi}_1^H\mathbf{h}_1$.

- The user quantizes the projected vector:

$$\hat{\mathbf{f}}_i = Q(\mathbf{f}_i) \in \mathcal{C}_i \subset \mathbb{C}^{d\times 1}, \tag{3.14}$$

where $Q(\cdot)$ is an quantizer, and $\mathcal{C}_i$ is a codebook for the ith block. Let B be the number of feedback bits such that $|\mathcal{C}_i| = 2^B$.

- The quantized vector $\hat{\mathbf{f}}_i$ is sent to the BS through a feedback link.
- Both of the BS and the user reconstruct the channel vector using the current feedback information and the previous reconstruction:

$$\hat{\mathbf{h}}_i = \eta \hat{\mathbf{h}}_{i-1} + \mathbf{\Psi}_i \hat{\mathbf{f}}_i \in \mathbb{C}^{M \times 1}. \tag{3.15}$$

With PBDF, we can reduce feedback overhead significantly as well as the encoding complexity since quantization is performed on the projected low dimensional subspace. Moreover, existing codebooks developed for small scale MIMO systems can be utilized. To reap such advantages while achieving the high throughput performance of massive MIMO systems, the projection matrix and the quantizer need to be properly designed exploiting the channel statistics.

The error vector $\mathbf{e}_i$ is defined as the difference between the original channel vector $\mathbf{h}_i$ and the reconstructed vector $\hat{\mathbf{h}}_i$, i.e.,

$$\begin{aligned} \mathbf{e}_i &= \mathbf{h}_i - \hat{\mathbf{h}}_i = \tilde{\mathbf{h}}_i - \mathbf{\Psi}_i \hat{\mathbf{f}}_i \\ &= \tilde{\mathbf{h}}_i - \mathbf{\Psi}_i \mathbf{f}_i + \mathbf{\Psi}_i (\mathbf{f}_i - \hat{\mathbf{f}}_i) = \mathbf{e}_{i,p} + \mathbf{\Psi}_i \mathbf{e}_{i,q}, \end{aligned} \tag{3.16}$$

where $\mathbf{e}_{i,p} = \tilde{\mathbf{h}}_i - \mathbf{\Psi}_i \mathbf{f}_i \in \mathbb{C}^{M \times 1}$ and $\mathbf{e}_{i,q} = \mathbf{f}_i - \hat{\mathbf{f}}_i \in \mathbb{C}^{d \times 1}$ are projection error and quantization error, respectively. Mean squared error (MSE), defined as

$$\text{MSE}_i = \mathbb{E}[\mathbf{e}_i^H \mathbf{e}_i], \tag{3.17}$$

can be decomposed as

$$\text{MSE}_i = \mathbb{E}[\mathbf{e}_{i,p}^H \mathbf{e}_{i,p}] + \mathbb{E}[\mathbf{e}_{i,q}^H \mathbf{e}_{i,q}], \tag{3.18}$$

due to the orthogonality between $\mathbf{e}_{i,p} = (\mathbf{I}_M - \mathbf{\Psi}_i \mathbf{\Psi}_i^H)\tilde{\mathbf{h}}_i$ and $\mathbf{\Psi}_i \mathbf{e}_{i,q}$. If we aim to minimize MSE, hence, the projection matrix $\mathbf{\Psi}_i$ and the quantizer $Q(\cdot)$ can be designed separately.

In the remainder of this section, the optimal projection matrix (in terms of MSE) is derived and combined with two representative quantization methods.

3.3.2 Projection for PBDF Framework

Projection into a smaller dimensional subspace is the key feature of the PBDF framework, and the choice of projection matrix has crucial impact on the system performance. In this subsection, the optimal projection matrix is derived in terms of MSE, which is closely related to received signal to noise ratio (SNR) in MIMO systems [73]. The validity of the MSE criterion is verified by numerical results in Section 3.5.

The optimal projector minimizing MSE is given by the well-known KLT [74], whose solution is the dominant eigenvectors of covariance matrix.

In the case of $i = 1$, where the original vector $\mathbf{h}_1$ is projected, its covariance matrix is $\mathbf{R}$ and the projection matrix minimizing MSE is given by the d dominant eigenvectors of $\mathbf{R}$, i.e.,

$$\begin{aligned}\mathbf{\Psi}_1 &= \underset{\mathbf{\Psi}}{\operatorname{argmin}}\, \mathbb{E}\left[\|\mathbf{h}_1 - \mathbf{\Psi}\mathbf{\Psi}^H\mathbf{h}_1\|_2^2\right] \\ &\overset{(a)}{=} \underset{\mathbf{\Psi}}{\operatorname{argmin}}\, \mathsf{tr}\left(\mathbf{R}\right) - \mathsf{tr}\left(\mathbf{\Psi}^H\mathbf{R}\mathbf{\Psi}\right) \\ &\overset{(b)}{=} \mathbf{U}_{[1:d]},\end{aligned} \tag{3.19}$$

where (a) comes from the trace invariance property under cyclic permutations, (b) is the result of KLT, and $\mathbf{U}_{[1:d]}$ is the first d columns of $\mathbf{U}$ defined in (3.9). The corresponding error covariance matrix is

$$\begin{aligned}\mathbf{E}_1 =& \mathbb{E}\left[\mathbf{e}_1\mathbf{e}_1^H\right] \\ =& (\mathbf{I}_M - \mathbf{\Psi}_1\mathbf{\Psi}_1^H)\mathbf{R}(\mathbf{I}_M - \mathbf{\Psi}_1\mathbf{\Psi}_1^H) + \mathbf{\Psi}_1\mathbf{E}_q\mathbf{\Psi}_1^H,\end{aligned} \tag{3.20}$$

where $\mathbf{E}_q = \mathbb{E}\left[\mathbf{e}_{1,q}\mathbf{e}_{1,q}^H\right]$ is the quantization error covariance matrix that depends on the quantizer.

For $i \geq 2$, since the difference vector $\tilde{\mathbf{h}}_i$ is projected, the projection matrix needs to adapt to its distribution. Following the correlation model (3.11), $\tilde{\mathbf{h}}_i$ can be decomposed

into previous error and the channel innovation, i.e.,

$$\begin{aligned}\tilde{\mathbf{h}}_i =&\mathbf{h}_i - \eta\hat{\mathbf{h}}_{i-1}\\ =&\underbrace{\eta(\hat{\mathbf{h}}_{i-1}+\mathbf{e}_{i-1})}_{\text{preserved}} + \underbrace{\sqrt{1-\eta^2}\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{g}_i}_{\text{innovative}} - \eta\hat{\mathbf{h}}_{i-1}\\ =&\eta\mathbf{e}_{i-1} + \sqrt{1-\eta^2}\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{g}_i,\end{aligned} \tag{3.21}$$

and difference covariance matrix is given by

$$\begin{aligned}\tilde{\mathbf{R}}_i =&\mathbb{E}\left[\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^H\right]\\ =&\eta^2\mathbb{E}\left[\mathbf{e}_{i-1}\mathbf{e}_{i-1}^H\right] + (1-\eta^2)\mathbf{R}\\ =&\eta^2\mathbf{E}_{i-1} + (1-\eta^2)\mathbf{R},\end{aligned} \tag{3.22}$$

where $\mathbf{e}_i$ is the error vector for fading block i defined in (3.16), and $\mathbf{E}_i = \mathbb{E}\left[\mathbf{e}_i\mathbf{e}_i^H\right]$ is its covariance matrix. By the same principle for the case of $i = 1$, the projection matrix is given by

$$\begin{aligned}\mathbf{\Psi}_i =&\underset{\mathbf{\Psi}}{\operatorname{argmin}}\,\mathbb{E}\left[\|\tilde{\mathbf{h}}_i - \mathbf{\Psi}\mathbf{\Psi}^H\tilde{\mathbf{h}}_i\|_2^2\right]\\ =&\underset{\mathbf{\Psi}}{\operatorname{argmin}}\,\mathbb{E}\left[\mathsf{tr}\left(\tilde{\mathbf{R}}_i\right) - \mathsf{tr}\left(\mathbf{\Psi}^H\tilde{\mathbf{R}}_i\mathbf{\Psi}\right)\right]\\ =&\tilde{\mathbf{U}}_{i,[1:d]},\end{aligned} \tag{3.23}$$

where $\tilde{\mathbf{R}}_i = \tilde{\mathbf{U}}_i\tilde{\mathbf{\Lambda}}_i\tilde{\mathbf{U}}_i^H$ is the eigendecomposition of $\tilde{\mathbf{R}}_i$ with $\tilde{\mathbf{\Lambda}}_i = \mathrm{diag}([\tilde{\lambda}_{i,1}^2, \tilde{\lambda}_{i,1}^2, \cdots, \tilde{\lambda}_{i,M}^2])$, such that $\tilde{\lambda}_{i,1} \geq \tilde{\lambda}_{i,2} \geq ... \geq \tilde{\lambda}_{i,M} \geq 0$.

Since $\mathbf{R}$ and η are known, the BS and the user can independently calculate the projection matrix if they can properly update $\mathbf{E}_{i-1}$. From the error expression of (3.16), error covariance can be expressed as

$$\begin{aligned}\mathbf{E}_i =&\mathbb{E}\left[\mathbf{e}_i\mathbf{e}_i^H\right]\\ =&\mathbb{E}\left[\mathbf{e}_{i,p}\mathbf{e}_{i,p}^H\right] + \mathbf{\Psi}_i\mathbb{E}\left[\mathbf{e}_{i,q}\mathbf{e}_{i,q}^H\right]\mathbf{\Psi}_i^H\\ =&(\mathbf{I}_M - \mathbf{\Psi}_i\mathbf{\Psi}_i^H)\tilde{\mathbf{R}}_i(\mathbf{I}_M - \mathbf{\Psi}_i\mathbf{\Psi}_i^H) + \mathbf{\Psi}_i\mathbf{E}_q\mathbf{\Psi}_i^H,\end{aligned} \tag{3.24}$$

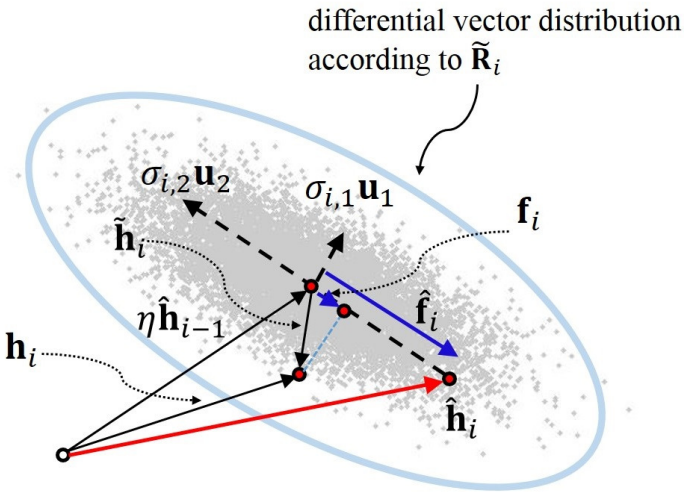

Figure 3.3: Illustration of projection-based differential feedback with $M = 2, d = 1$, and a real valued channel.

where the former part is the projection error determined by difference covariance $\tilde{\mathbf{R}}_i$ and projection matrix $\mathbf{\Psi}_i$, and

$$\mathbf{E}_q = \mathbb{E}\left[\mathbf{e}_{i,q}\mathbf{e}_{i,q}^H\right] \tag{3.25}$$

in the latter part is the quantization error depending on the quantizer. Note that $\mathbf{E}_q$ is independent to i since it is an expectation over distribution.

By updating $\tilde{\mathbf{R}}_i$ using (3.22) and $\mathbf{E}_i$ using (3.24) iteratively, the BS and the user can calculate the projection matrix. Note that the update process only uses the channel statistics, and does not depend on channel realization. Hence, the BS and the user can calculate the sequence of $\mathbf{\Psi}_i$ a priori and use it for feedback process until channel statistics change.

Fig. 3.3 illustrates the concept of the PBDF process with a simple example of $M = 2$ and $d = 1$, and a pseudo code of projection matrix calculation is shown in Algorithm 1.

Algorithm 1 Calculation of projection matrix

Input $\mathbf{R}, \eta, \mathbf{E}_q$

▷ Channel statistics and quantization error statistics

Output $\{\mathbf{\Psi}_i | i = 1, 2, \cdots\}$

Initialization

$\tilde{\mathbf{R}} \leftarrow \mathbf{R}$ ▷ Difference covariance $\tilde{\mathbf{R}} = \mathbb{E}\left[\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H\right]$

$\mathbf{E} \leftarrow \mathbf{0}_{M \times M}$ ▷ Error covariance $\mathbf{E} = \mathbb{E}\left[\mathbf{e}\mathbf{e}^H\right]$

$i \leftarrow 1$ ▷ Block index i

while $\mathbf{R}$ remains constant **do**

$\tilde{\mathbf{U}}, \tilde{\mathbf{\Lambda}} \leftarrow \text{EVD}(\tilde{\mathbf{R}})$ ▷ Eigendecomposition of $\tilde{\mathbf{R}}$ such that $\tilde{\mathbf{R}} = \tilde{\mathbf{U}}, \tilde{\mathbf{\Lambda}}\tilde{\mathbf{U}}^H$, $\tilde{\mathbf{U}} = [\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2, ..., \tilde{\mathbf{u}}_M]$, $\tilde{\mathbf{\Lambda}} = \text{diag}([\tilde{\lambda}_1^2, \tilde{\lambda}_2^2, ..., \tilde{\lambda}_M^2])$, and $\tilde{\lambda}_1^2 \geq \tilde{\lambda}_2^2 \geq ... \geq \tilde{\lambda}_M^2$

$\mathbf{\Psi} \leftarrow [\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2, ..., \tilde{\mathbf{u}}_d]$ ▷ Projection matrix $\mathbf{\Psi}$

$\mathbf{E} \leftarrow (\mathbf{I}_M - \mathbf{\Psi}\mathbf{\Psi}^H)\tilde{\mathbf{R}}(\mathbf{I}_M - \mathbf{\Psi}\mathbf{\Psi}^H) + \mathbf{\Psi}\mathbf{E}_q\mathbf{\Psi}^H$

▷ Update of error covariance $\mathbf{E}$

$\tilde{\mathbf{R}} \leftarrow \eta^2\mathbf{E} + (1 - \eta^2)\mathbf{R}$

▷ Update of difference covariance $\tilde{\mathbf{R}}$

return $\mathbf{\Psi}_i \leftarrow \mathbf{\Psi}$ ▷ Projection matrix

$i \leftarrow i + 1$

end while

The process of updating the error covariance matrix $\mathbf{E}_i$ in (3.24) requires the quantization error covariance $\mathbf{E}_q$, which depends on the choice of quantizer. In other words, since the error performance depends on the quantization method, the quantization effect needs to be considered when calculating projection matrix. In this chapter, two representative quantization methods including scalar quantization (SQ) and random vector quantization (RVQ) are applied for the PBDF framework, and corresponding error covariance matrices ($\mathbf{E}_q$) are derived.

Scalar quantization

SQ is the simplest way to quantize the projected low dimensional vector $\mathbf{f}_i$. I assume that only one bit is allocated to each real/imaginary part of elements to avoid bit-allocation problem. Consequently, its application is restricted to the case when the number of feedback bit is the twice of the dimension of projection space, i.e. $B = 2d$.

For a Gaussian random variable v with variance σ^2, the optimal quantizer minimizing MSE is given by $Q_{\text{SQ}}(v) = \sigma\sqrt{\frac{2}{\pi}}\text{sign}(v)$, and corresponding expected quantization error is $\mathbb{E}\left[|v - Q_{\text{SQ}}(v)|^2\right] = (1 - \frac{2}{\pi})\sigma^2$ [75]. Hence, when the SQ is applied to PBDF, quantizer and error covariance are given by

$$\hat{\mathbf{f}}_{\text{SQ},i} = \sqrt{\frac{\tilde{\mathbf{\Lambda}}_{i,[1:d]}}{\pi}}\left[\text{sign}(\text{Re}\{\mathbf{f}_i\}) + j\text{sign}(\text{Im}\{\mathbf{f}_i\})\right], \tag{3.26}$$

and

$$\mathbf{E}_{q,\text{SQ}} = (1 - \frac{2}{\pi})\tilde{\mathbf{\Lambda}}_{i,[1:d]}, \tag{3.27}$$

where $\tilde{\mathbf{\Lambda}}_{i,[1:d]}$ is diagonal matrix whose elements are d largest eigenvalues of $\tilde{\mathbf{R}}_i$. Note that since the BS has the knowledge of $\tilde{\mathbf{\Lambda}}_{i,[1:d]}$, only B bits of feedback (sign information) are required for reconstruction.

Random vector quantization

To quantize the projected vector more efficiently, vector quantization (VQ) approach can be used. For the simplicity in codebook generation and error covariance calculation, RVQ is considered in this chapter. Let $\mathcal{C} = \{\mathbf{c}_b \in \mathbb{C}^{d\times 1} | b = 1, ..., 2^B\}$ be a randomly generated codebook such that $\|\mathbf{c}_b\|_2^2 = 1$ for all b. Since the codebook is distributed isotropically, we first normalize $\mathbf{f}_i$ to be isotropic, i.e., $\bar{\mathbf{f}}_i = \tilde{\mathbf{\Lambda}}_{i,[1:d]}^{-\frac{1}{2}} \mathbf{f}_i$. Then the user computes the quantization index $b^\star$ according to

$$b^\star = \underset{b=1,...,2^B}{\operatorname{argmax}} |\mathbf{c}_b^H \bar{\mathbf{f}}_i|, \tag{3.28}$$

and feeds back $b^\star$ and $\alpha = \mathbf{c}_{b^\star}^H \bar{\mathbf{f}}_i$ together. Note that α is also fed back for the reconstruction process, which is distinguished from conventional (non-differential) VQ schemes where only the index is fed back.

After receiving feedback, the BS performs reconstruction according to

$$\hat{\mathbf{f}}_{\text{RVQ},i} = \tilde{\mathbf{\Lambda}}_{i,[1:d]}^{\frac{1}{2}} (\alpha \mathbf{c}_{b^\star}). \tag{3.29}$$

Since $\bar{\mathbf{f}}_i - \alpha \mathbf{c}_{b^\star}$ is isotropically distributed and

$$\begin{aligned} \mathbb{E}\left[\|\bar{\mathbf{f}}_i - \alpha \mathbf{c}_{b^\star}\|_2^2\right] =& \mathbb{E}\left[\|\bar{\mathbf{f}}_i\|_2^2 - |\alpha|^2\right] \\ =& d - \mathbb{E}\left[|\alpha|^2\right], \end{aligned} \tag{3.30}$$

error covariance matrix is given by [59]

$$\begin{aligned} \mathbf{E}_{q,\text{RVQ}} =& \mathbb{E}\left[(\mathbf{f}_i - \hat{\mathbf{f}}_i)(\mathbf{f}_i - \hat{\mathbf{f}}_i)^H\right] \\ =& \tilde{\mathbf{\Lambda}}_{i,[1:d]}^{\frac{1}{2}} \mathbb{E}\left[(\bar{\mathbf{f}}_i - \alpha \mathbf{c}_{b^\star})(\bar{\mathbf{f}}_i - \alpha \mathbf{c}_{b^\star})^H\right] \tilde{\mathbf{\Lambda}}_{i,[1:d]}^{\frac{1}{2}} \\ =& \frac{d - \mathbb{E}\left[|\alpha|^2\right]}{d} \tilde{\mathbf{\Lambda}}_{i,[1:d]}, \end{aligned} \tag{3.31}$$

where $\mathbb{E}\left[|\alpha|^2\right] = d \sum_{b=0}^{2^B} \frac{\binom{2^B}{b}^{(-1)^b}}{b(d-1)+1}$.

3.3.3 Efficient Algorithm

The procedure finding the optimal projection matrix for each channel block includes the eigendecomposition of an $M \times M$ matrix $\tilde{\mathbf{R}}_i$. Since the eigendecomposition requires $\mathcal{O}(M^3)$ multiplications, it becomes a huge burden in massive MIMO systems with very large M.

However, if $\tilde{\mathbf{R}}_i$ inherits eigenstructure from $\mathbf{R}$, the calculation of projection matrix can be simplified. Note that columns of the first projection matrix $\mathbf{\Psi}_1$ are part of the eigenvectors of $\mathbf{R}$. Following (3.24), hence, $\mathbf{E}_1$ has same eigenvectors with $\mathbf{R}$ if the quantization error covariance $\mathbf{E}_q$ is diagonal as for the two aforementioned quantizers (SQ and RVQ). Then, by (3.22), $\tilde{\mathbf{R}}_2$ has the same eigenvectors and columns of $\mathbf{\Psi}_2$ are also d eigenvectors of $\mathbf{R}$. In the same manner, we can prove that the eigenvectors of $\tilde{\mathbf{R}}_i$ for any $i \geq 2$ are still eigenvectors of $\mathbf{R}$. Consequently, we can find the $\mathbf{\Psi}_i$ without eigendecomposition at every block. Instead, the eigendecomposition is performed once at the beginning and the selection of d columns out of M columns of $\mathbf{U}$ is needed at each block. Using this observation, $\tilde{\mathbf{R}}_i$ can be rewritten as

$$\tilde{\mathbf{R}}_i = \mathbf{U}\mathbf{\Sigma}_i\mathbf{U}^H, \tag{3.32}$$

where $\mathbf{\Sigma}_i = \text{diag}([\sigma_{i,1}^2, \sigma_{i,2}^2, ..., \sigma_{i,M}^2])$ and $\sigma_{i,m}^2$ is the eigenvalue corresponding to $\mathbf{u}_m$. Note that $[\sigma_{i,1}^2, \sigma_{i,2}^2, ..., \sigma_{i,M}^2]$ is a reordered version of $[\tilde{\lambda}_{i,1}^2, \tilde{\lambda}_{i,2}^2, ..., \tilde{\lambda}_{i,M}^2]$. While $\tilde{\lambda}_{i,m}^2$ are ordered as $\tilde{\lambda}_{i,1} \geq \tilde{\lambda}_{i,2} \geq ... \geq \tilde{\lambda}_{i,M} \geq 0$, $\sigma_{i,m}^2$ are reordered such that the eigenvector matrix is identical to the $\mathbf{U}$ (eigenvector matrix of $\mathbf{R}$).

With this expression, the update process of (3.22) and (3.24) can be replaced by

$$\sigma_{i,m}^2 = (1-\eta^2)\lambda_m^2 + \eta^2\sigma_{i-1,m}^2\mathbf{u}_m^H\mathbf{\Psi}_{i-1}\mathbf{E}_q\mathbf{\Psi}_{i-1}^H\mathbf{u}_m, \tag{3.33}$$

if $\mathbf{u}_m$ is a column of $\mathbf{\Psi}_{i-1}$ and

$$\sigma_{i,m}^2 = \eta^2\sigma_{i-1,m}^2 + (1-\eta^2)\lambda_m^2, \tag{3.34}$$

otherwise. After updating $[\sigma_{i,1}^2, \sigma_{i,2}^2, ..., \sigma_{i,M}^2]$, $\mathbf{\Psi}_i$ is constructed by collecting the $\mathbf{u}_m$ corresponding to the d largest $\sigma_{i,m}^2$.

Note that only necessary condition for this simplification is diagonal $\mathbf{E}_q$. The efficient algorithm with only one eigenvalue decomposition is described in Algorithm 2 in the next page.

3.4 Discussions

In this section, I address some practical issues on implementing the proposed feedback scheme in massive MIMO systems.

3.4.1 Projection with Imperfect CSIR

While I have assumed that the user knows corresponding channel exactly, downlink channel estimation is also a challenging issue in massive MIMO systems. Since downlink resource proportional to the number of transmit antennas is required for the estimation, it is nearly impossible to achieve perfect CSIR in practical massive MIMO systems. Hence, the imperfection of CSIR needs to be considered in feedback process.

To solve the channel estimation problem, several recent works have aimed to reduce the length of training sequence by exploiting spatial and temporal correlation of the channel [20], [19]. According to the works, second order statistics of channel estimation error can be obtained if the channel statistics and pilot signal are given. Based on the estimation error statistics, the projection matrix can be modified for environment where downlink channel training is imperfect. In this subsection, I extend the PBDF framework to imperfect CSIR scenarios and develop *distortion-aware* PBDF, which incorporates the effect of estimation error in projection matrix calculation.

Algorithm 2 Efficient algorithm calculating projection matrix

Input $\mathbf{R}, \eta,$ diagonal $\mathbf{E}_q$ ▷ Channel statistics and quantization error statistics

Output $\{\mathbf{\Psi}_i | i = 1, 2, \cdots\}$

Initialization

$\mathbf{U}, \mathbf{\Lambda} \leftarrow \text{EVD}(\mathbf{R})$ ▷ Eigendecomposition of $\mathbf{R}$ such that $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_M]$, $\mathbf{\Lambda} = \text{diag}([\lambda_1^2, \lambda_2^2 ..., \lambda_M^2])$, and $\lambda_1^2 \geq \lambda_2^2 \geq ... \geq \lambda_M^2$

for $m = 1$ to M **do**

$\sigma_m^2 \leftarrow \lambda_m^2$

end for

$S \leftarrow \emptyset$ ▷ The set of indices of selected eigenvectors

$i \leftarrow 1$ ▷ Block index i

while $\mathbf{R}$ remains constant **do**

$S \leftarrow$ indices corresponding to the d largest σ_m^2

$\mathbf{\Psi} \leftarrow [\mathbf{u}_{s_1} \mathbf{u}_{s_2} ... \mathbf{u}_{s_d}]$, where $S = \{s_1, s_2, ..., s_d\}$

for $m = 1$ to M **do**

if $m \in S$ **then**

$\sigma_m^2 \leftarrow \eta^2 \sigma_m^2 \mathbf{u}_m^H \mathbf{\Psi} \mathbf{E}_q \mathbf{\Psi}^H \mathbf{u}_m + (1 - \eta^2) \lambda_m^2$

else

$\sigma_m^2 \leftarrow \eta^2 \sigma_m^2 + (1 - \eta^2) \lambda_m^2$

end if

end for

return $\mathbf{\Psi}_i \leftarrow \mathbf{\Psi}$ ▷ Projection matrix

$i \leftarrow i + 1$

end while

Let $\bar{\mathbf{h}}_i$ and $\bar{\mathbf{e}}_i$ denote estimated channel and estimation error at block i, such that

$$\mathbf{h}_i = \bar{\mathbf{h}}_i + \bar{\mathbf{e}}_i, \tag{3.35}$$

and assume that covariance matrix of the estimation error

$$\bar{\mathbf{E}}_i = \mathbb{E}\left[\bar{\mathbf{e}}_i \bar{\mathbf{e}}_i^H\right] \tag{3.36}$$

is known to both of the BS and the user.

For $i = 1$, in this condition, the optimal projection matrix can be obtained by solving

$$\begin{aligned}
\mathbf{\Psi}_1 =& \underset{\mathbf{\Psi}}{\operatorname{argmin}} \mathbb{E}\left[\|\mathbf{h}_1 - \mathbf{\Psi}\mathbf{\Psi}^H \bar{\mathbf{h}}_1\|_2^2\right] \\
=& \underset{\mathbf{\Psi}}{\operatorname{argmin}} \mathbb{E}\left[\|\mathbf{h}_1 - \mathbf{\Psi}\mathbf{\Psi}^H (\mathbf{h}_1 - \bar{\mathbf{e}}_1)\|_2^2\right] \\
=& \underset{\mathbf{\Psi}}{\operatorname{argmin}} \operatorname{tr}(\mathbf{R}) - \operatorname{tr}\left(\mathbf{\Psi}^H (\mathbf{R} - \bar{\mathbf{E}}_1)\mathbf{\Psi}\right),
\end{aligned} \tag{3.37}$$

whose solution is the d dominant eigenvectors of $\mathbf{R} - \bar{\mathbf{E}}_1$, and corresponding error covariance matrix is given by

$$\begin{aligned}
\mathbf{E}_1 =& \mathbb{E}\left[\mathbf{e}_1 \mathbf{e}_1^H\right] \\
=& \mathbf{R} - \mathbf{\Psi}_1 \mathbf{\Psi}_1^H \left(\mathbf{R} - \bar{\mathbf{E}}_1\right) \mathbf{\Psi}_1 \mathbf{\Psi}_1^H + \mathbf{\Psi}_1 \mathbf{E}_q \mathbf{\Psi}_1^H.
\end{aligned} \tag{3.38}$$

For $i \geq 2$, similarly to the case of $i = 1$, the objective function of (3.23) can be modified into

$$\begin{aligned}
\mathbf{\Psi}_i =& \underset{\mathbf{\Psi}}{\operatorname{argmin}} \mathbb{E}\left[\|\tilde{\mathbf{h}}_i - \mathbf{\Psi}\mathbf{\Psi}^H \left(\bar{\mathbf{h}}_i - \eta \hat{\mathbf{h}}_{i-1}\right)\|_2^2\right] \\
=& \underset{\mathbf{\Psi}}{\operatorname{argmin}} \mathbb{E}\left[\|\tilde{\mathbf{h}}_i - \mathbf{\Psi}\mathbf{\Psi}^H \left(\tilde{\mathbf{h}}_i - \bar{\mathbf{e}}_i\right)\|_2^2\right] \\
=& \underset{\mathbf{\Psi}}{\operatorname{argmin}} \operatorname{tr}\left(\tilde{\mathbf{R}}_i\right) - \operatorname{tr}\left(\mathbf{\Psi}^H (\tilde{\mathbf{R}}_i - \bar{\mathbf{E}}_i)\mathbf{\Psi}\right),
\end{aligned} \tag{3.39}$$

where $\tilde{\mathbf{R}}_i = \mathbb{E}\left[\tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^H\right] = \mathbb{E}\left[(\mathbf{h}_i - \eta\hat{\mathbf{h}}_{i-1})(\mathbf{h}_i - \eta\hat{\mathbf{h}}_{i-1})^H\right]$ is updated according to

(3.22). As a result, the solution is the d dominant eigenvectors of $\tilde{\mathbf{R}}_i - \bar{\mathbf{E}}_i$ and corresponding error covariance matrix is resulted as

$$\begin{aligned}\mathbf{E}_i =& \mathbb{E}\left[\mathbf{e}_i \mathbf{e}_i^H\right] \\ =& \tilde{\mathbf{R}}_i - \mathbf{\Psi}_i \mathbf{\Psi}_i^H \left(\tilde{\mathbf{R}} - \bar{\mathbf{E}}_i\right) \mathbf{\Psi}_i \mathbf{\Psi}_i^H + \mathbf{\Psi}_i \mathbf{E}_q \mathbf{\Psi}_i^H .\end{aligned} \tag{3.40}$$

Consequently, projection matrix in PBDF framework can adapt to erroneous CSIR by incorporating the estimation error in the projection matrix calculation process. Moreover, if the estimation error covariance $\bar{\mathbf{E}}_i$ is also diagonalizable with $\mathbf{U}$ as in [19], the simplified calculation of Algorithm 2 can be applied.

3.4.2 Acquisition of Channel Statistics

To implement the PBDF, the BS as well as the user needs to have knowledge on the channel statistics $\mathbf{R}$ and η. Since these values change in a much slower rate than the channel response [76], long-term feedback can be a solution. Moreover, there exist some alternative methods which reduce the burden further.

There are rich literature on downlink channel covariance estimation [77]- [79]. One way is to use uplink direction-of-arrival (DOA) information which can be estimated with subspace-based techniques [77]. More efficiently, the BS can estimate spatial covariance matrix $\mathbf{R}$ using the uplink covariance directly [78], or with a frequency calibration [79].

Since temporal correlation coefficient η depends on the mobility of the user [37], the estimate of uplink temporal correlation can be used with calibration. The estimation of uplink correlation coefficient falls into the auto-regressive parameter estimation problem which has been studied in [80] and [81].

In multi-carrier systems such as OFDM, channel statistics $\mathbf{R}$ and η are approximately identical across subcarriers, even if the channel responses are frequency-selective

[82]. Hence, the same projection matrix $\mathbf{\Psi}_i$ can be used for feedback of different subcarriers, making the effort for channel statistics acquisition more negligible.

3.5 Simulation Results

In this section, the effectiveness of the proposed feedback scheme is numerically verified thorough simulations performed in three deferent channel models. To see the performance in massive MIMO systems, I assume a BS with $M = 64$ antennas and users each with a single antenna. At every fading block, the channel is quantized in $B = 8$ bits and delivered to the BS through error-free feedback link. For PBDF, the difference vector is projected into a $d = 4$ dimensional subspace. The perfect knowledge of channel statistics ($\mathbf{R}$ and η) at both of the transmit and the receiver sides is assumed, and both of the perfect CSIR and the imperfect CSIR due to training error are considered. For temporal correlation, Jakes' model is adopted, i.e., $\eta = J_0(2\pi f_D T_B)$ where $J_0(\cdot)$, $f_D = \frac{v f_c}{c}$, and T_B denote the 0th order Bessel function of the first kind, the maximum Doppler frequency, and the block period, respectively. I set $\eta = 0.9924$ by assuming the user speed $v = 3$ km/h, carrier frequency $f_c = 2$ GHz, $c = 3 \times 10^8$ m/s and $T_B = 5$ ms.

The proposed scheme is compared to three different feedback schemes. RVQ is a basic codebook whose codewords are isotropically distributed and neither spatial nor temporal correlation is exploited [59]. Rotation codebook is generated by rotating the RVQ codebook to the direction of spatial correlation matrix [62]. Polar-cap differential codebook generates their codewords on a spherical cap centered around the previous quantization [65]. Since I consider spatially and temporally correlated channels, I use a modified version of polar-cap differential codebook that incorporates the discrete Fourier transform (DFT) codebook, which was also presented in [65].

In Fig. 3.4, expected normalized inner product between the original and recon-

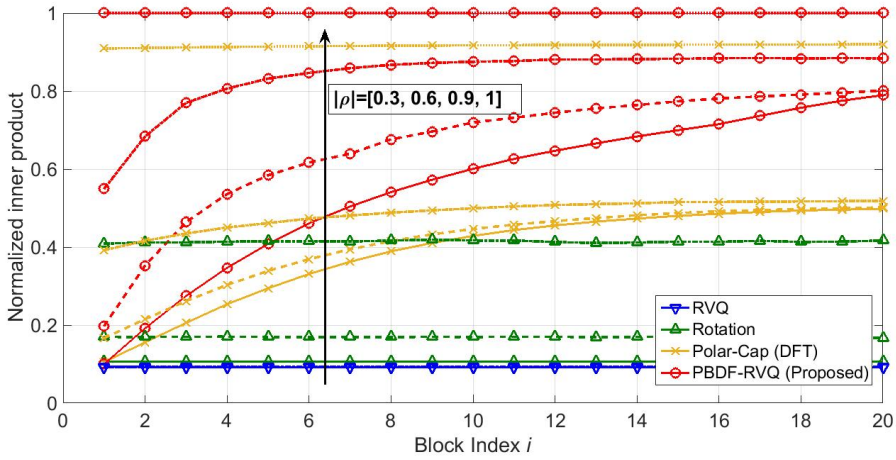

Figure 3.4: Normalized inner product versus fading block index with different spatial correlation levels.

structed channel vectors, $\mathbb{E}\left[\frac{|\mathbf{h}_i^H \hat{\mathbf{h}}_i|}{\|\mathbf{h}_i\|_2 \|\hat{\mathbf{h}}_i\|_2}\right]$, of the proposed scheme is compared to various feedback schemes with different spatial correlation levels. To analyze the effect of spatial correlation, exponential correlation model is firstly considered. In the exponential model, the channel covariance matrix is given by

$$[\mathbf{R}]_{m,n} = \begin{cases} \rho^{|m-n|}, & \text{if} \quad m \leq n \\ \left(\rho^{|m-n|}\right)^*, & \text{otherwise,} \end{cases} \tag{3.41}$$

where $\rho = |\rho| e^{j\theta}$ ($|\rho| \in [0,1], \theta \in [0, 2\pi]$) is correlation coefficient between adjacent antennas. Normalized inner products when $|\rho| \in [0.3, 0.6, 0.9, 1]$ and θ is uniformly distributed are plotted. As expected, the performance of the proposed scheme is better when spatial correlation is higher. With all correlation levels, PBDF outperforms other schemes by exploiting the knowledge of $\mathbf{R}$ and η jointly.

In Fig. 3.5, normalized inner product of different feedback schemes are compared with different temporal correlation levels. Spatial covariance matrix was generated according to the exponential model with $\rho = 0.9$, while other conditions are set iden-

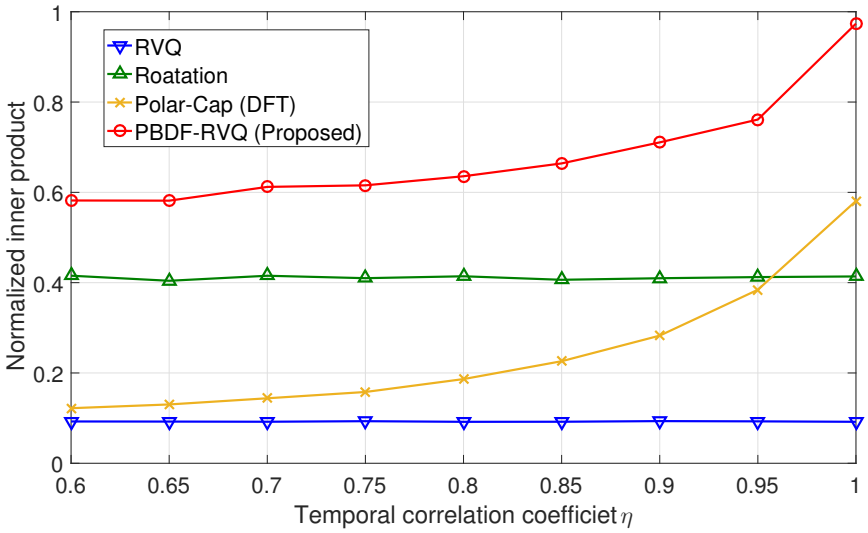

Figure 3.5: Normalized inner product versus temporal correlation.

tically to Fig. 3.4. Since the proposed scheme reconstructs channels using previous feedback information, the feedback accuracy of the porposed scheme depends on the level of temporal correlation. From this result, we can expect that the proposed scheme is more effective for slow-fading channels

Fig. 3.6 and Fig. 3.7 show the normalized inner products of various feedback schemes when a realistic *one-ring model* [83] with uniform linear array (ULA) is adopted as a spatial correlation model. The channel covariance matrix of ULA one-ring model is given by

$$[\mathbf{R}]_{m,n} = \frac{1}{2\Delta} \int_{-\Delta+\theta}^{\Delta+\theta} e^{jD(m-n)\sin(\alpha)} d\alpha, \tag{3.42}$$

where Δ and θ are angular spread (AS) and angle of arrival (AoA), respectively. AoA and AS are set to $\theta = 0$ and $\Delta = \frac{\pi}{10}$. In Fig. 3.6, perfect CSIR is assumed, while erroneous CSIR is assumed in Fig. 3.7 considering the difficulty of channel training in massive MIMO systems. In Fig. 3.7, the CSIR was obtained using the scheme pro-

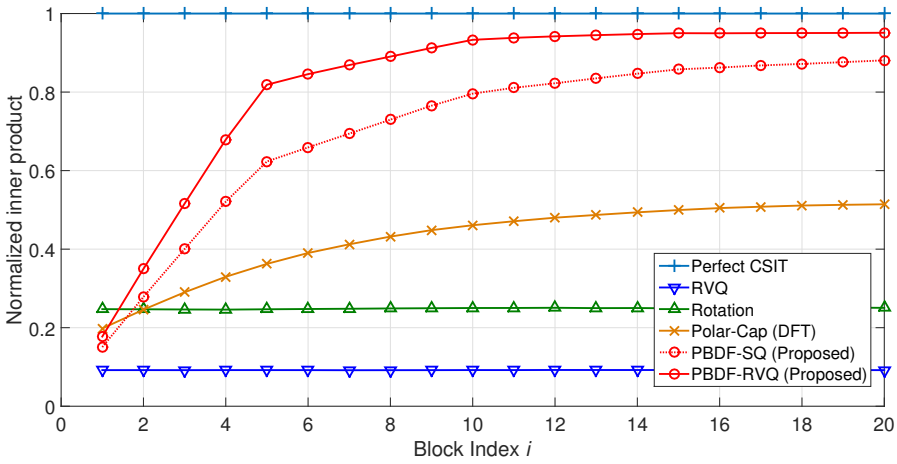

Figure 3.6: Normalized inner product versus fading block index with ULA antenna deployment and perfect CSIR.

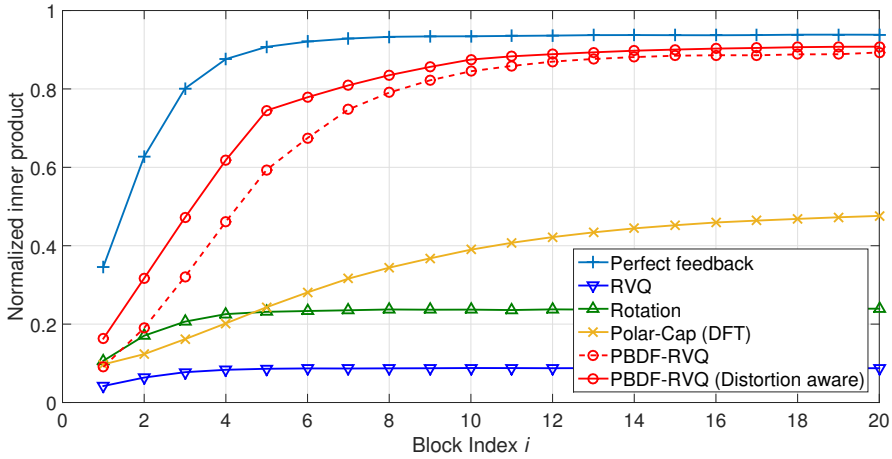

Figure 3.7: Normalized inner product versus fading block index with ULA antenna deployment and imperfect CSIR.

posed in [20], where the transmitter sends a random orthogonal training sequence of length $T_p = 8$ every block, and the receiver estimates the channel by a Kalman filter. In both cases, normalized inner product close to the perfect feedback case can be attained

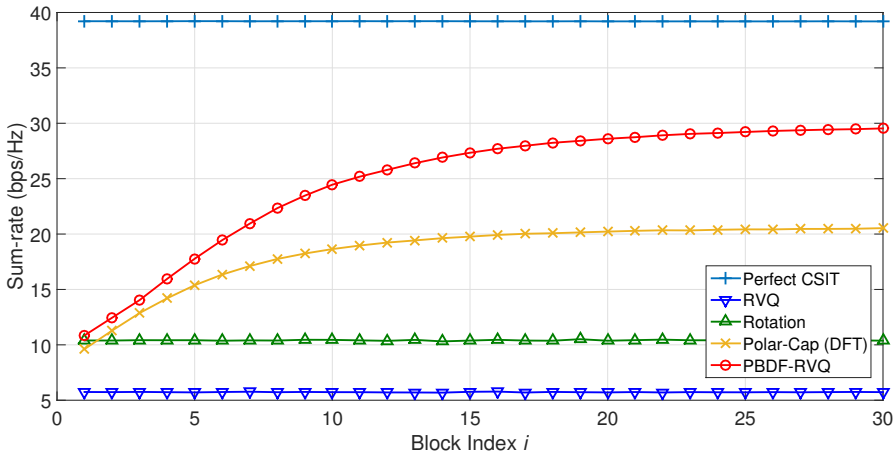

Figure 3.8: Sum-rate performances versus fading block index with ULA antenna deployment and perfect CSIR.

with PBDF even though the feedback scheme is designed to minimize MSE after convergence. By exploiting spatial and temporal correlation simultaneously, moreover, the proposed scheme outperforms other schemes. It should be noted that there is a performance gap between the proposed scheme and the polar-cap codebook which also utilizes temporal and spatial correlation by generating codewords nearby the past codeword, and by adopting DFT codebook. The result implies that explicit exploitation of the channel statistics is desirable to achieve accurate CSIT when the number of feedback bits is extremely small compared to the number of antennas. Moreover, the performance of the distortion-aware PBDF, developed in Sec. 3.4.1, shows better performance when the CSIR is imperfect since it calculates the optimal projection considering the distortion from the training process. The impact of the distortion-awareness is more noticeable at earlier fading blocks, when the estimation error is more significant.

The sum-rate of the ULA system is depicted in Fig. 3.8 and Fig. 3.9, assuming perfect and imperfect CSIR conditions, respectively. ZFBF was adopted as the mul-

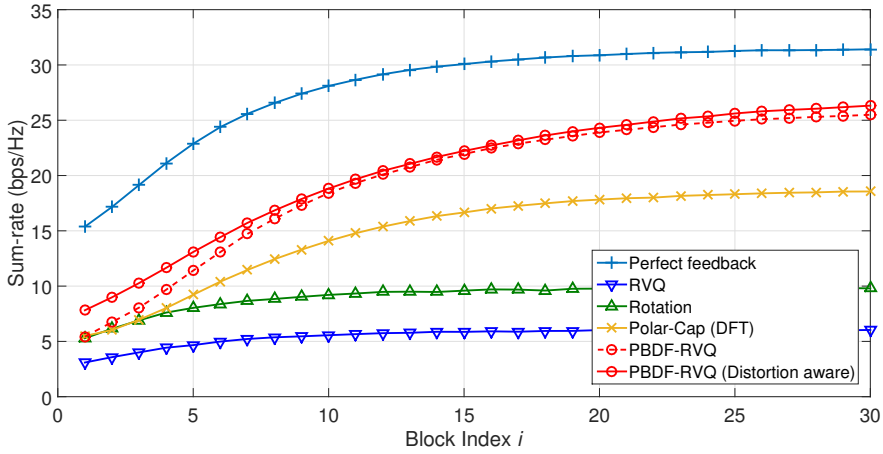

Figure 3.9: Sum-rate performances versus fading block index with ULA antenna deployment and imperfect CSIR.

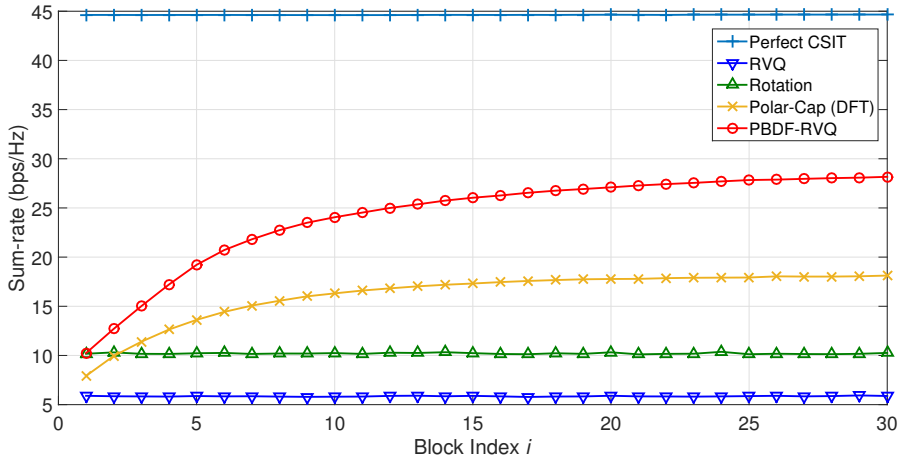

Figure 3.10: Sum-rate performances versus fading block index with UPA antenna deployment and perfect CSIR.

tiuser precoding strategy, and the achievable sum-rate was obtained through (3.5) and (3.6). On the assumption of ULA one-ring model, AoA and AS for $K = 6$ users are generated from uniform distribution of $\theta_k \in [-\pi/6, \pi/6]$ and $\Delta_k \in [\pi/12, \pi/6]$.

Table 3.1: Comparison of encoding complexity

	RVQ	Polar-cap	PBDF-SQ	PBDF-RVQ
Number of complex multiplications	$2^B M$	$2^B(M^2+M)$	dM	$dM + 2^B d$

The transmit power P is set to 10 dB. By using PBDF combined with RVQ, we can achieve a large portion of the expected sum-rate which is achievable with perfect feedback. In Fig. 3.10, the sum-rate performances of various schemes are shown with the assumption of $(M_v \times M_h) = (8 \times 16)$ uniform planar array (UPA) antennas, which are deployed to support beamforming in horizontal and vertical domains. According to the results of [56], the spatial covariance matrix of the UPA deployment was obtained through the Kronecker product of the two ULA covariance matrices, i.e.,

$$\mathbf{R} = \mathbf{R}_v \otimes \mathbf{R}_h, \tag{3.43}$$

where $\mathbf{R}_v$ and $\mathbf{R}_h$ denote the vertical and horizontal covariance matrices which are calculated as (3.42). The vertical AS and AoA were set to $\Delta_v = \pi/10$ and $\theta_v = \pi/6$, and the horizontal AS and AoA were randomly drawn from the same distribution as for the ULA setting. Similarly to the ULA scenario, the proposed feedback scheme shows better sum-rate performance than other schemes with a UPA deployment.

The lower error of the proposed scheme is achieved with a lower encoding complexity compared to the conventional schemes. When RVQ or rotation codebook is used, the encoding of each block channel requires $2^B M$ complex multiplications to find the most aligned codeword according to (7). The polar-cap codebook, which rotates the codebook every block interval, requires $2^B M^2$ more complex multiplications for the rotation in addition to $2^B M$ multiplications for encoding. With the proposed scheme, on the other hand, the number of multiplications for encoding is dM (if SQ is applied) or $dM + 2^B d$ (if VQ is applied) since the encoding is performed on the

reduced dimension. Here, the computation for projection matrix is not taken into account since the sequence of matrices can be calculated a priori. With $M = 64$, $d = 4$, and $B = 8$, the proposed scheme with VQ requires only 8% (1.6% for the SQ case) of the multiplication complexity compared to the conventional schemes such as RVQ or rotational codebook. The complexity comparison is summarized in Table 3.1 in the next page.

3.6 Conclusion

In this chapter, I proposed a novel feedback framework for FDD massive MIMO systems. The proposed feedback scheme, referred to as projection-based differential feedback (PBDF), can efficiently reduce the amount of feedback by combining the concepts of differential feedback and dimension reduction. Exploiting spatial and temporal correlation simultaneously, the optimal projection matrix for PBDF framework was also derived along with a low-complexity algorithm. Simulation results showed that the proposed feedback scheme can accomplish accurate CSIT with low encoding complexity, even when the number of feedback bits is small.

Chapter 4

Mitigating Pilot Contamination via Pilot Design

In massive multiple-input multiple-output (MIMO) systems, the effect of inter-user interference and noise can be suppressed through simple signal processing techniques. However, *pilot contamination* effect, which results from the use of a correlated pilot sequence, cannot be eliminated by increasing the number of antennas. Therefore, pilot contamination is a performance bottleneck for massive MIMO networks. In this chapter, the effect of the pilot contamination is analyzed, and two approaches for alleviating the pilot contamination are proposed for multi-cell massive MIMO systems. First, a pilot design algorithm based on alternating minimization is developed. Second, a scalable channel training framework, referred as to *generalized pilot reuse scheme*, is presented, and it is shown that Grassmannian subspace packing provides an optimal sequence set for the framework. Contrary to existing pilot reuse protocols that restrict the pilot length to be an integer multiple of the number of users, the proposed methods leverage a pilot with an arbitrary length. Hence, pilot sequence can be designed more flexibly to maximize spectral efficiency.

4.1 Introduction

Massive multiple-input multiple-output (MIMO) systems are expected to provide unprecedented spectral and energy efficiency for fifth generation (5G) cellular networks [1, 2]. By deploying a large number of antennas at base stations (BS), the spectral and energy efficiency can increase without bound [5, 6, 8]. Moreover, these advantages can be achieved by using simple transmission and reception techniques, such as maximal ratio transmission (MRT) and maximal ration combining (MRC).

However, some practical issues need to be addressed to realize such benefits [4, 9]. For example, it is very important to acquire the channel state information (CSI) since the aforementioned benefits of massive MIMO systems are results of coherent transmission/reception that exploits accurate CSI. At the BS, the uplink CSI is acquired from predefined pilot sequences transmitted by the user equipments (UEs) during the training period and is used to detect the uplink signals.[1] When the pilot sequences for different UEs are correlated, the estimated CSI of a UE is corrupted by the CSI of other UEs, and as a result, inter-user interference cannot be suppressed. This detrimental effect, which is known as *pilot contamination*, has been pointed out as a bottleneck to achieve a high spectral efficiency in massive MIMO systems [4, 5, 9].

Due to its importance, much interest has been directed at mitigating the pilot contamination effect, and various techniques have been proposed. In [84], the use of time-shifted pilot sequences was proposed to avoid pilot contamination. However, the gain is suspicious since the channel estimation is corrupted by downlink signals from different cells, which are generally transmitted with a higher power than an uplink pilot. In [85], a pilot assignment strategy was proposed to assign the same pilot sequence to UEs with approximately orthogonal channel covariances. However, it is not trivial

[1] In time division duplexing (TDD) systems, the uplink CSI is also used for downlink transmission exploiting the channel reciprocity.

to acquire and share the channel covariance. A practical approach, which assigns orthogonal sequences to neighboring cells, was proposed and investigated in [86]. The scheme, which is referred to as pilot reuse, has been modified in various ways. Fractional pilot reuse, which assigns orthogonal sequence only to UEs in the cell edge, was proposed in [87], and an adaptive pilot assignment scheme was proposed for non-symmetric networks in [88]. In such schemes, however, the pilot sequences of different UEs are restricted to be same or orthogonal. As a result, the length of the sequence is limited to be the product of the pilot reuse factor and the number of users in a cell. Although the optimal reuse factor was investigated in [90], very few options are available for optimization.

In this chapter, the use of more generalized pilot sequence is investigated to allow the pilot sequence to have an arbitrary length. Also, two approaches for training with an arbitrary length pilot is proposed. Firstly, for a given sequence length, I propose a pilot design algorithm that minimizes the correlation between the pilot sequences that are used by different cells. Secondly, a scalable training framework, which is referred to as *generalized pilot reuse*, is presented, and it is shown that Grassmannian subspace packing provides an optimal sequence set for the framework. The proposed schemes alleviate the pilot contamination and provide more flexible optimization opportunity for uplink training to maximize the spectral efficiency of the network.

4.2 System Model

The uplink transmission of a multi-cell multi-user massive MIMO system is considered. The system consists of L cells, and $\mathcal{L} = \{1, \cdots, L\}$ is the set of cell indices. Each cell is composed of a BS with M antennas and K single-antenna UEs that transmit data streams to the corresponding BS. Since a massive MIMO system is considered, I assume $K \ll M$.

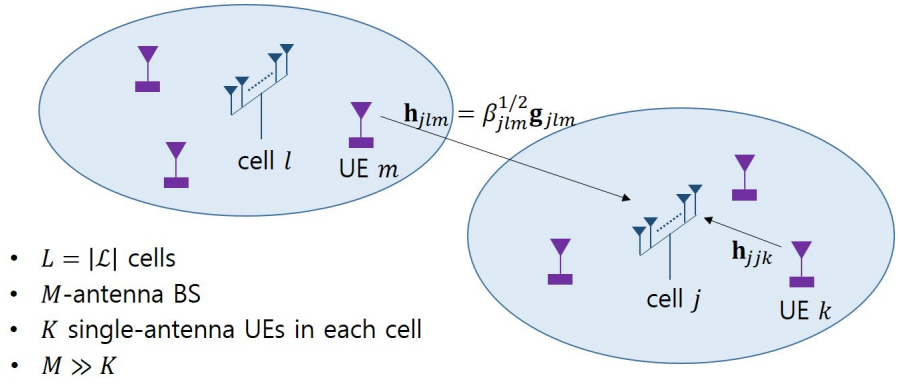

Figure 4.1: A multi-cell multi-user massive MIMO system.

4.2.1 Multi-cell Massive MIMO Systems

The $M \times 1$ received vector of the BS in cell j is

$$\begin{aligned}\mathbf{y}_j =& \sum_{l\in\mathcal{L}}\sum_{k=1}^{K}\sqrt{\rho_u}\mathbf{h}_{jlk}x_{lk} + \mathbf{n}_j \\ =& \sqrt{\rho_u}\sum_{l\in\mathcal{L}}\mathbf{H}_{jl}\mathbf{x}_l + \mathbf{n}_j,\end{aligned} \tag{4.1}$$

where ρ_u is the signal-to-noise ratio (SNR) of the uplink, x_{lk} is the transmit symbol of UE k in cell l with the constraint $\mathbb{E}\{|x_{lk}|^2\} = 1$, $\mathbf{h}_{jlk}$ is the $M \times 1$ channel vector between UE k in cell l and the BS in cell j, and $\mathbf{n}_j$ denotes an additive noise vector, which is distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$. In the matrix form, $\mathbf{x}_l = [x_{l1}, \cdots, x_{lK}]^T$ is the vector of the transmit symbols, and $\mathbf{H}_{jl} = [\mathbf{h}_{jl1}, \cdots \mathbf{h}_{jlK}]$ is the $M \times K$ channel matrix between the K UEs in cell l and the BS in cell j. The uplink model of a multi-cell multi-user system is depicted in Fig. 4.1.

The channel vector $\mathbf{h}_{jlk}$ can be written as

$$\mathbf{h}_{jlk} = \beta_{jlk}^{1/2}\mathbf{g}_{jlk}, \tag{4.2}$$

where $\mathbf{g}_{jlk}$ is a vector for fast fading, which is distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$, and β_{jlk} models the attenuation effect and shadow fading. Accordingly, β_{jlk} can be expressed

as

$$\beta_{jlk} = \frac{z_{jlk}}{d_{jlk}^{\kappa}}, \tag{4.3}$$

where d_{jlk} is the distance between UE k in cell l and the BS in cell j, κ is the path-loss exponent, and z_{jlk} is the shadow fading coefficient which follows the log-normal distribution, i.e., $10\log(z_{jlk})$ follows the zero-mean Gaussian distribution with a variance of σ_S^2. In the same way, the channel matrix $\mathbf{H}_{jl}$ can be rewritten as

$$\mathbf{H}_{jl} = \mathbf{G}_{jl}\mathbf{D}_{jl}^{1/2}, \tag{4.4}$$

where $\mathbf{G}_{jl} = [\mathbf{g}_{jl1}, \cdots, \mathbf{g}_{jlK}]$, and $\mathbf{D}_{jl}$ is a $K \times K$ diagonal matrix whose elements are given as $[\mathbf{D}_{jl}]_{k,k} = \beta_{jlk}$.

In this chapter, I focus on *asymptotically massive MIMO systems* where M increases without bounds. As $M \to \infty$, the law of large numbers (LLN) kicks in [5], i.e.,

$$\frac{1}{M}\mathbf{h}_{jlk}^H \mathbf{n}_j \to 0, \tag{4.5}$$

and

$$\frac{1}{M}\mathbf{g}_{jlk}^H \mathbf{g}_{j'l'k'} \to \delta_{jj'}\delta_{ll'}\delta_{kk'}, \tag{4.6}$$

where $\delta_{ii'}$ denotes the Kronecker delta. As a result, we have

$$\frac{1}{M}\mathbf{H}_{jl}^H \mathbf{H}_{j'l'} \to \mathbf{D}_{jl}\delta_{jj'}\delta_{ll'}. \tag{4.7}$$

4.2.2 Uplink Channel Training

I assume a block transmission where a part of a block is allocated for channel training and the remaining part is utilized for data transmission. Let B channel uses constitute a transmission block, and assume that the channel response is constant over a block. If orthogonal frequency division multiplexing (OFDM) is utilized, the block size B is designed such that $B \leq T_c W_c$, where the channel response is constant over T_c OFDM

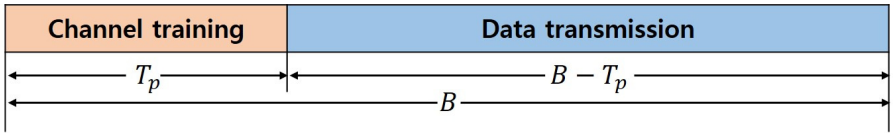

Figure 4.2: Block structure with training and data transmission phases.

symbol times and W_c subcarriers. For T_p channel uses, the UEs transmit predefined pilot sequences for channel training, and the BS estimates the channel responses. During the data transmission phase, the UEs transmit data symbols and the BS detects the symbols transmitted by the corresponding UEs using the estimated channel. The block structure is depicted in Fig. 4.2.

During the training phase of T_p channel uses, the BS in cell j receives

$$\mathbf{Y}_{T,j} = \sum_{l \in \mathcal{L}} \sum_{k=1}^{K} \sqrt{\rho_u} \mathbf{h}_{jlk} \boldsymbol{\phi}_{lk}^H + \mathbf{N}_{T,j}, \tag{4.8}$$

where $\boldsymbol{\phi}_{lk} \in \mathbb{C}^{T_p \times 1}$ is a pilot sequence transmitted by UE k in cell l and $\mathbf{N}_{T,j}$ is a noise matrix whose elements are independent and identically distributed (i.i.d.) complex Gaussian random variables (RVs) with unit variance. I assume that $T_p \geq K$ and all UEs in a same cell use orthogonal pilots, i.e., $\boldsymbol{\phi}_{lk}^H \boldsymbol{\phi}_{lk'} = \delta_{kk'} T_p$. In a matrix form, the orthogonality constraint can be written as

$$\boldsymbol{\Phi}_l^H \boldsymbol{\Phi}_l = T_p \mathbf{I}_K, \quad \forall l \in \mathcal{L}, \tag{4.9}$$

where $\boldsymbol{\Phi}_l = [\boldsymbol{\phi}_{l1}, \cdots, \boldsymbol{\phi}_{lK}]$. With the received signal, the BS estimates the channel responses between the UEs in cell j and itself, i.e., $\{\mathbf{h}_{jjk}\}$. As a simplest way, the estimate of $\mathbf{h}_{jjk}$ can be obtained by correlating the received training signal $\mathbf{Y}_{T,j}$ with pilot sequence $\boldsymbol{\phi}_{jk}$, i.e.,

$$\begin{aligned} \hat{\mathbf{h}}_{jjk} &= \frac{1}{T_p \sqrt{\rho_u}} \mathbf{Y}_{T,j} \boldsymbol{\phi}_{jk} \\ &= \mathbf{h}_{jjk} + \frac{1}{T_p} \sum_{l \in \mathcal{L} \setminus \{j\}} \sum_{m=1}^{K} \mathbf{h}_{jlm} \boldsymbol{\phi}_{lm}^H \boldsymbol{\phi}_{jk} + \frac{1}{T_p \sqrt{\rho_u}} \bar{\mathbf{n}}_{T,j}, \end{aligned} \tag{4.10}$$

where $\bar{\mathbf{n}}_{T,j} = \mathbf{N}_{T,j}\boldsymbol{\phi}_{jk}$. Note that $\bar{\mathbf{n}}_{T,j}$ is distributed as $\mathcal{CN}(\mathbf{0}, T_p\mathbf{I}_M)$ since the elements of $\mathbf{N}_{T,j}$ are i.i.d. complex Gaussian RVs.

4.2.3 Data Transmission

During the transmission phase, the UEs transmit their data symbols and the BS detects the transmitted symbols from the received signal. I assume an MRC receiver as the receiving filter for the simplicity. With the MRC receiver, the BS estimates the symbol transmitted by UE k by correlating the estimated channel vector with the received vector, i.e.,

$$\hat{\mathbf{x}}_{jk} = \hat{\mathbf{h}}_{jjk}^H \mathbf{y}_j. \tag{4.11}$$

Leveraging the results of the LLN in (4.5)-(4.7), we have

$$\frac{1}{M}\hat{\mathbf{x}}_{jk} \xrightarrow[M\to\infty]{} \sqrt{\rho_u}\beta_{jjk}x_{jk} + \frac{\sqrt{\rho_u}}{T_p}\sum_{l\in\mathcal{L}\setminus\{j\}}\sum_{m=1}^{K}\boldsymbol{\phi}_{jk}^H\boldsymbol{\phi}_{lm}\beta_{jlm}x_{lm}. \tag{4.12}$$

This result shows that by increasing the number of the BS antennas without bounds, the effect of the noise and intra-cell interference vanishes, and the desired signal and inter-cell interference caused by the correlated pilots remain.

As a result, the signal-to-interference ratio of UE k in cell j becomes

$$SIR_{jk} = \frac{T_p^2\beta_{jjk}^2}{\sum_{l\in\mathcal{L}\setminus\{j\}}\sum_{m=1}^{K}|\boldsymbol{\phi}_{jk}^H\boldsymbol{\phi}_{lm}|^2\beta_{jlm}^2}, \tag{4.13}$$

and the ergodic achievable rate of the UE is bounded as

$$\begin{aligned} R_{jk} =& \mathbb{E}_{\{\beta\}}\left[\log_2\left(1+SIR_{jk}\right)\right] \\ \geq& \log_2\left(1+\frac{T_p^2}{\mathbb{E}_{\{\beta\}}\left[\sum_{l\in\mathcal{L}\setminus\{j\}}\sum_{m=1}^{K}\frac{\beta_{jlm}^2}{\beta_{jjk}^2}|\boldsymbol{\phi}_{jk}^H\boldsymbol{\phi}_{lm}|^2\right]}\right) \\ =& \log_2\left(1+\frac{T_p^2}{\sum_{l\in\mathcal{L}\setminus\{j\}}\mu_j^l\sum_{m=1}^{K}|\boldsymbol{\phi}_{jk}^H\boldsymbol{\phi}_{lm}|^2}\right), \end{aligned} \tag{4.14}$$

where μ_j^l is defined as

$$\mu_j^l = \mathbb{E}_{\{\beta\}}\left[\frac{\beta_{jlm}^2}{\beta_{jjk}^2}\right], \tag{4.15}$$

and the inequality comes from Jensen's inequality applied for a convex function $f(x) = \log_2(1+\frac{1}{x})$. Note that μ_j^l is defined independently to UE index since it is averaged over the distribution of UE location as well as the shadow fading. Moreover, since μ_j^l is determined by the BS location and the distribution of UE location, I assume that the values do not change and that these are known during the pilot design process.

Using the above result, the achievable sum-rate of all UEs in cell j can be expressed as

$$\begin{aligned} R_j^{\text{sum}} &= \sum_{k=1}^{K} R_{jk} \\ &\geq K \log_2\left(1 + \frac{T_p^2}{\frac{1}{K}\sum_{l\in\mathcal{L}\setminus\{j\}} \mu_l^l \sum_{k=1}^{K}\sum_{m=1}^{K} |\boldsymbol{\phi}_{jk}^H \boldsymbol{\phi}_{lm}|^2}\right), \end{aligned} \tag{4.16}$$

where the inequality also comes from Jensen's inequality. The network-wide sum-rate, defined as the sum of the achievable rate of all UEs in the network, is also bounded as

$$\begin{aligned} R^{\text{sum}} &= \sum_{j\in\mathcal{L}} R_j^{\text{sum}} \\ &\geq LK \log_2\left(1 + \frac{T_p^2}{\frac{1}{LK}\sum_{j\in\mathcal{L}}\sum_{l\in\mathcal{L}\setminus\{j\}} \mu_j^l \|\boldsymbol{\Phi}_j^H \boldsymbol{\Phi}_l\|_F^2}\right), \end{aligned} \tag{4.17}$$

where $\|\mathbf{A}\|_F$ is the Frobenius norm of matrix $\mathbf{A}$, i.e., $\|\mathbf{A}\|_F = \sqrt{\mathsf{tr}\left(\mathbf{A}\mathbf{A}^H\right)}$. Incorporating the training overhead, the net spectral efficiency is given as

$$R_{\text{net}}^{\text{sum}} = \frac{B - T_p}{B} R^{\text{sum}}. \tag{4.18}$$

4.3 Iterative Pilot Design Algorithm

The achievable rate analysis in Section 4.2 shows that as the number of BS antennas increases, the intra-cell interference and noise are averaged out, so the inter-cell in-

terference caused by the correlated pilot becomes the only performance bottleneck. Unfortunately, the degradation resulting from the pilot correlation, which is referred to as *pilot contamination effect*, cannot be eliminated by increasing the transmit power or the number of antennas. To eliminate pilot contamination, orthogonal pilot sequences need to be used by the UEs in all cells. However, a perfect orthogonalization of the sequences, which requires $T_p \geq KL$ channel uses, would consume extensive resources, eventually degrading the net spectral efficiency since the size of a block is limited by the coherence time and frequency. In this section, I propose an algorithm, which designs a set of pilot sequences to minimize the pilot contamination effect with a given pilot sequence length T_p.

4.3.1 Algorithm

The lower bound of the network throughput in (4.17) can be maximized by minimizing the weighted sum of the correlations between the pilot matrices of different cells. Thus, I propose a pilot design algorithm that minimizes the correlation between the pilot sequences that are used by the different cells, i.e., the optimization problem is given as

$$\begin{aligned} \underset{\{\mathbf{\Phi}_i | i \in \mathcal{L}\}}{\text{minimize}} \quad & \sum_{j \in \mathcal{L}} \sum_{l \in \mathcal{L} \setminus \{j\}} \mu_j^l \|\mathbf{\Phi}_j^H \mathbf{\Phi}_l\|_F^2 \\ \text{subject to} \quad & \mathbf{\Phi}_i^H \mathbf{\Phi}_i = T_p \mathbf{I}_K, \quad \forall i \in \mathcal{L}. \end{aligned} \tag{4.19}$$

Note that the objective function is invariant to the rotation of the matrices, i.e., for any $j \in \mathcal{L}$, the replacement of $\mathbf{\Phi}_j$ by $\mathbf{\Phi}_j \mathbf{U}$ makes no difference in the objective function value if $\mathbf{U} \in \mathbb{C}^{K \times K}$ is a unitary matrix.

Due to the orthogonality constraint, the problem is not convex. To solve the optimization problem, I propose an iterative algorithm, which updates the pilot sequence of a cell in each iteration. With respect to a typical cell b, the objective function can be

decomposed as

$$\begin{aligned}&\sum_{j\in\mathcal{L}}\sum_{l\in\mathcal{L}\setminus\{j\}}\mu_j^l\|\mathbf{\Phi}_j^H\mathbf{\Phi}_l\|_F^2\\&\quad=\sum_{l\in\mathcal{L}\setminus\{b\}}(\mu_l^b+\mu_b^l)\|\mathbf{\Phi}_b^H\mathbf{\Phi}_l\|_F^2+\sum_{j\in\mathcal{L}\setminus\{b\}}\sum_{l\in\mathcal{L}\setminus\{b,j\}}\|\mathbf{\Phi}_j^H\mathbf{\Phi}_l\|_F^2,\end{aligned}\tag{4.20}$$

where the first summation is a function of $\mathbf{\Phi}_b$ while the second summation is independent to the choice of $\mathbf{\Phi}_b$. Hence, when the other $L-1$ pilot sequences are fixed, the pilot sequence of cell b minimizing the objective function is given as

$$\begin{aligned}\mathbf{\Phi}_b^\star=&\operatorname*{argmin}_{\mathbf{\Phi}_b}\sum_{l\in\mathcal{L}\setminus\{b\}}(\mu_l^b+\mu_b^l)\|\mathbf{\Phi}_b^H\mathbf{\Phi}_l\|_F^2\\=&\sqrt{T_p}\nu_{min}^K\left(\sum_{l\in\mathcal{L}\setminus\{b\}}\left(\mu_l^b+\mu_b^l\right)\mathbf{\Phi}_l\mathbf{\Phi}_l^H\right),\end{aligned}\tag{4.21}$$

where $\nu_{min}^K(\mathbf{A})$ is a $T_p\times K$ matrix, whose columns are the K least dominant eigenvectors of $\mathbf{A}\in\mathbb{C}^{T_p\times T_p}$. With the proposed algorithm, a pilot sequence set $\{\mathbf{\Phi}_i\}$ is designed by iteratively updating the pilot sequences until the objective function converges. The proposed pilot design procedure is summarized in Alg. 3.

Algorithm 3 Pilot Sequence Design

1: **Input** $\{\mu_l^j|l\in\mathcal{L},j\in\mathcal{L}\}$
2: **Initialize** Generate arbitrary pilot sequence $\{\mathbf{\Phi}_i|i\in\mathcal{L}\}$ satisfying $\mathbf{\Phi}_i\mathbf{\Phi}_i^H=T_p\mathbf{I}_K,\forall i\in\mathcal{L}$.
3: **while** not converged **do**
4: **for** $j=1:L$ **do**
5: $\mathbf{\Phi}_j=\sqrt{T_p}\nu_{min}^K\left(\sum_{l\in\mathcal{L}\setminus\{j\}}\left(\mu_l^j+\mu_j^l\right)\mathbf{\Phi}_l^H\mathbf{\Phi}_l\right)$
6: **end for**
7: **end while**
8: **Output** $\{\mathbf{\Phi}_i|i\in\mathcal{L}\}$

4.3.2 Proof of Convergence

Although the proposed algorithm is not guaranteed to find a global optimal solution, the convergence of the algorithm can be intuitively proved. First, note that the objective function is lower bounded by zero. Also, every update of the pilot sequence reduces the value of the objective function. Hence, the value of the objective function must converge to some value that is not less than zero.

4.4 Generalized Pilot Reuse

While the scheme proposed in Section 4.3 provides an opportunity to design a pilot sequence of an arbitrary length minimizing the pilot contamination, the algorithm requires information of long-term statistics on interference power (4.15). Moreover, since it relies on the iterative updates of each cell's pilot, it is hard to apply the scheme to large networks with a large number of cells. In this section, a scalable channel training framework, referred as to *generalized pilot reuse scheme*, is presented to mitigate the pilot contamination in large cellular networks. Moreover, it is shown that Grassmannian subspace packing can be exploited to find an optimal sequence set.

4.4.1 Concept of Pilot Reuse Schemes

Pilot contamination has been pointed as a performance bottleneck since the seminal analysis of massive MIMO [5]. Due to its importance, various directions to mitigating the pilot contamination have been explored. Among them, pilot reuse was proposed as a practical approach [86]. Inspired by the frequency reuse of conventional cellular networks, the pilot reuse scheme assigns orthogonal sequences to adjacent cells exploiting the observation that the most of the interfering power comes from neighboring cells. Various schemes have been also developed by modifying the pilot reuse

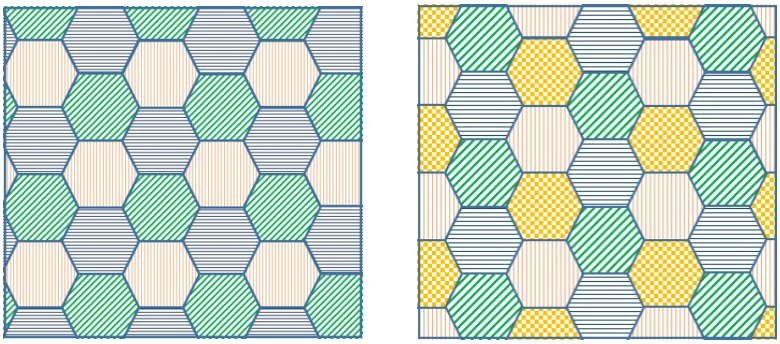

Figure 4.3: Pilot reuse patterns with reuse factor $G = 3$ (left) and $G = 4$ (right).

scheme, including fractional pilot reuse, which assigns orthogonal sequence only to UEs in the cell edge [87], and an adaptive pilot assignment scheme, which was developed for non-symmetric networks [88]. In such schemes, however, the pilot sequences of different UEs are restricted to be same or orthogonal. When pilot reuse scheme is used with a reuse factor G, G different sets of pilot are utilized in order that adjacent cells do not use the same pilot. The concept of the pilot reuse is illustrated in Fig. 4.3. In the figure, the cells using the same pilot sequence are represented by the same color.

4.4.2 Pilot Design based on Grassmannian Subspace Packing

I consider a system using generalized pilot reuse scheme with a reuse factor G. Then, cells in the network is divided into G groups such that the cells in the same group shares the same pilot sequences. Let $\mathbf{\Phi}_g = [\boldsymbol{\phi}_g^{(1)}, ..., \boldsymbol{\phi}_g^{(K)}] \in \mathbb{C}^{T_p \times K}$ be the training matrix for group $g \in \mathcal{G} = \{1, 2, ..., G\}$. Then, the estimated channel (4.10) becomes

$$\begin{aligned}\hat{\mathbf{h}}_{jjk} &= \frac{1}{T_p\sqrt{\rho_u}}\mathbf{Y}_{T,j}\boldsymbol{\phi}_{g(j)}^{(k)} \\ &= \mathbf{h}_{jjk} + \sum_{l \in \mathcal{L}_{g(j)} \setminus \{j\}} \mathbf{h}_{jlk} + \frac{1}{T_p} \sum_{l \in \mathcal{L} \setminus \mathcal{L}_{g(j)}} \sum_{m=1}^{K} \mathbf{h}_{jlm} (\boldsymbol{\phi}_{g(l)}^{(m)})^H \boldsymbol{\phi}_{g(j)}^{(k)} + \frac{1}{T_p\sqrt{\rho_u}} \bar{\mathbf{n}}_{T,j},\end{aligned} \tag{4.22}$$

where $g(j)$ and $\mathcal{L}_g$ denote the index of the group including cell j and the set of cells in group g, respectively. Then, the maximal ratio combining of (4.12) also can be expressed in a similar way, and the asymptotic SIR is given as

$$SIR_{jk} \to \frac{T_p^2 \beta_{jjk}^2}{T_p^2 \sum_{l \in \mathcal{L}_{g(j)} \setminus \{j\}} \beta_{jlk}^2 + \sum_{l \in \mathcal{L} \setminus \mathcal{L}_{g(j)}} \sum_{m=1}^{K} \beta_{jlm}^2 |(\boldsymbol{\phi}_{g(l)}^{(m)})^H \boldsymbol{\phi}_{g(j)}^{(k)}|^2}. \quad (4.23)$$

Then, after a procedure similar to Section 4.3, we can derive a per-cell achievable sum-rate as

$$R^{\text{sum}} \geq K \log_2 \left(1 + \frac{T_p^2}{A + B \frac{1}{GK} \sum_{g=1}^{G} \sum_{g' \neq g} \|\boldsymbol{\Phi}_g^H \boldsymbol{\Phi}_{g'}\|_F^2} \right), \quad (4.24)$$

where

$$A = \mathbb{E}_{\{\beta\}}[\sum_{l \in \mathcal{L}_{g(j)} \setminus \{j\}} \frac{\beta_{jlm}^2}{\beta_{jjk}^2}], \quad (4.25)$$

and

$$B = \mathbb{E}_{\{\beta\}}[\sum_{l \in \mathcal{L}_{g'}} \frac{\beta_{jlm}^2}{\beta_{jjk}^2}], \quad \text{for } g' \neq g. \quad (4.26)$$

Then, pilot design problem becomes

$$\begin{aligned} \underset{\{\boldsymbol{\Phi}_g | g \in \mathcal{G}\}}{\text{minimize}} \quad & \sum_{g=1}^{G} \sum_{g' \neq g} \|\boldsymbol{\Phi}_g^H \boldsymbol{\Phi}_{g'}\|_F^2 \\ \text{subject to} \quad & \boldsymbol{\Phi}_g^H \boldsymbol{\Phi}_g = T_p \mathbf{I}_K, \quad \forall g \in \mathcal{G}. \end{aligned} \quad (4.27)$$

Note that the problem is different with (4.19) in that the weight terms are eliminated thanks to the symmetry in group geometry. For the simplified objective function, we can derive a lower bound as stated in the following theorem.

Theorem 1. For a set of G matrices satisfying $\boldsymbol{\Phi}_g^H \boldsymbol{\Phi}_g = T_p \mathbf{I}_K$, for all g, the objective function in (4.27) is lower bounded as

$$\sum_{g=1}^{G} \sum_{g' \neq g} \|\boldsymbol{\Phi}_g^H \boldsymbol{\Phi}_{g'}\|_F^2 \geq (GK - T_p) GK T_p. \quad (4.28)$$

Proof. The objective function can be rewritten as below using the definition of Frobenius norm and trace properties:

$$\begin{aligned}
\sum_{g=1}^{G}\sum_{g'\neq g}\|\mathbf{\Phi}_g^H\mathbf{\Phi}_{g'}\|_F^2 =& \sum_{g=1}^{G}\sum_{g'=1}^{G}\|\mathbf{\Phi}_g^H\mathbf{\Phi}_{g'}\|_F^2 - \sum_{g=1}^{G}\|\mathbf{\Phi}_g^H\mathbf{\Phi}_g\|_F^2 \\
=& \sum_{g=1}^{G}\sum_{g'=1}^{G}\mathsf{tr}(\mathbf{\Phi}_g^H\mathbf{\Phi}_{g'}\mathbf{\Phi}_{g'}^H\mathbf{\Phi}_g) - \sum_{g=1}^{G}\mathsf{tr}(\mathbf{\Phi}_g^H\mathbf{\Phi}_g\mathbf{\Phi}_g^H\mathbf{\Phi}_g) \\
=& \sum_{g=1}^{G}\sum_{g'=1}^{G}\mathsf{tr}(\mathbf{\Phi}_g\mathbf{\Phi}_g^H\mathbf{\Phi}_{g'}\mathbf{\Phi}_{g'}^H) - GKT_p^2 \\
=& \mathsf{tr}(\sum_{g=1}^{G}\sum_{g'=1}^{G}\mathbf{\Phi}_g\mathbf{\Phi}_g^H\mathbf{\Phi}_{g'}\mathbf{\Phi}_{g'}^H) - GKT_p^2 \\
=& \mathsf{tr}[(\sum_{g=1}^{G}\mathbf{\Phi}_g\mathbf{\Phi}_g^H)(\sum_{g'=1}^{G}\mathbf{\Phi}_{g'}\mathbf{\Phi}_{g'}^H)] - GKT_p^2.
\end{aligned} \tag{4.29}$$

Let define a matrix $\mathbf{S} = \sum_{g=1}^{G}\mathbf{\Phi}_g\mathbf{\Phi}_g^H$, and its eigenvalue decomposition as $\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, where $\mathbf{U}$ is the eigenvector matrix and $\mathbf{\Lambda} = \mathrm{diag}[\lambda_1, ..., \lambda_{T_p}]$ is a diagonal matrix whose elements are the eigenvalues. Then, the lower bound of Theorem 1 can be derived from the result of (4.29) as belows:

$$\begin{aligned}
\sum_{g=1}^{G}\sum_{g'\neq g}\|\mathbf{\Phi}_g^H\mathbf{\Phi}_{g'}\|_F^2 =& \mathsf{tr}(\mathbf{S}^2) - GKT_p^2 \\
=& \sum_{i=1}^{T_p}\lambda_i^2 - GKT_p^2 \\
\overset{(a)}{\geq}& \frac{1}{T_p}(\sum_{i=1}^{T_p}\lambda_i)^2 - GKT_p^2 \\
=& \frac{1}{T_p}(\mathsf{tr}(\mathbf{S}))^2 - GKT_p^2 \\
=& G^2K^2T_p - GKT_p^2,
\end{aligned} \tag{4.30}$$

where (a) comes form the Cauchy's inequality. □

Moreover, the lower bound presented in Theorem 1 can be attained by well known Grassmannian subspace packing [92]. The result in [92] implies that $\max_{g \neq g'} \|\mathbf{\Phi}_g^H \mathbf{\Phi}_{g'}\|_F^2 = \frac{(GK-T_p)KT_p}{G-1}$ can be achieved through the Grassmannian subspace packing. Hence, we can claim that the equality of (4.28) is attainable by packing G K-dimensional subspaces into a T_p-dimensional subspace.

4.5 Simulation Results

4.5.1 Iterative Pilot Design

In simulations, a network with $L = 3$ cells was considered with the regular hexagonal layout depicted in Fig. 4.4. In the network, the BSs are located at the center of the corresponding cells with a radius r (m), and in each cell, $K = 10$ UE locations were generated according to the uniform distribution over the cell area, with the exception of a circle of r_h m around each BS. The channel was realized according to (4.2) and (4.3) with a path-loss exponent $\kappa = 3$ and shadow fading with a standard deviation $\sigma_S = 8$ dB. SNR was set as $\rho_u = 10$ dB on average, and a channel block with $B = 200$ was considered, assuming that the channel is constant over $T_c = 20$ OFDM symbols and $W_c = 10$ subcarriers. The parameters are summarized in Table 4.1. For the algorithm, statistical information μ_l^j for all $j \neq l$ were obtained using a simulation, which randomly generated the locations and large scale fading coefficients for 10^6 UEs for each cell.

Fig. 4.5 shows the per-cell net sum-rate when the length of the pilot varies from K to LK. Note that the corner points $T_p = K$ and $T_p = LK$ denote the full pilot reuse (the same pilot) and the no pilot reuse (orthogonal pilots), respectively. In all cases, the pilot designed by the proposed algorithm (blue lines) provides higher spectral efficiency than random orthogonal pilot (orange lines). The green lines indicate

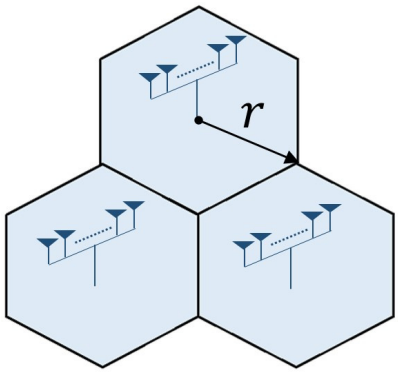

Figure 4.4: A network of three regular hexagonal cells.

Table 4.1: Simulation parameters

cell radius	$r = 1600$ m
hole radius	$r_h = 0.2r$
path-loss exponent	$\kappa = 3$
shadow fading deviation	$\sigma_S = 8$ dB
average SNR	$\rho_u = 10$ dB
block size	$B = 200$

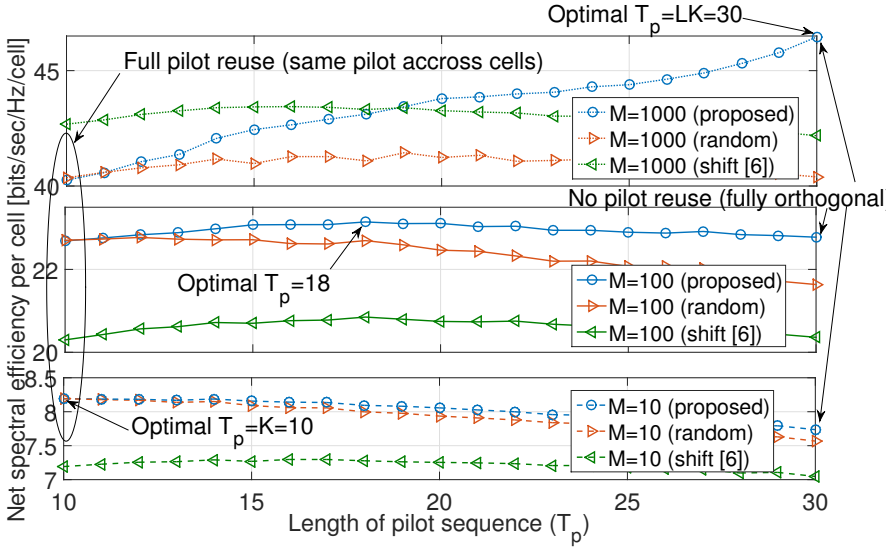

Figure 4.5: Net spectral efficiency per cell.

the sum-rate when pilot shift scheme of [84] is used. According to the specs of practical cellular systems, downlink transmission power was set to be 20 dB larger than uplink power. While the pilot contamination can be mitigated with the pilot shift, the channel estimation is corrupted by the strong downlink signal of adjacent cells, which

eventually degrades the performance.

As more channel uses are allocated for training, the correlation between the pilot sequences becomes smaller, while the resource available for data transmission decreases. In a system with a small number of antennas (e.g. $M = 10$), small T_p allowing a large fraction of a block to the data transmission is favorable in terms of the spectral efficiency since the effect of noise and intra-cell interference remains significant. On the other hand, with an extremely large number of antennas (e.g. $M = 1000$), the effect of noise and intra-cell interference almost vanishes, hence orthogonalizing the pilot with a large T_p is beneficial. In a reasonable massive MIMO regime (10s to 100s antennas), the length of the training sequence maximizing spectral efficiency is between K and LK and depends on the number of antennas and the coherence block length.

Note that either the reuse of the same pilot across cells ($T_p = K$) or perfect orthogonalization ($T_p = LK$), the only possible options with conventional pilot reuse schemes [46, 90], cannot provide the optimal spectral efficiency when the number of antennas is reasonably large. However, with the proposed scheme, which allows a pilot design with an arbitrary length, more flexible design of pilot sequence that maximizes the spectral efficiency is viable.

4.5.2 Generalized Pilot Reuse

For the generalized pilot reuse, I conducted a simulation similar to the previous section. Unlike the previous section, which considered a small network, a cellular network with a large number of hexagonal cells is considered. By means of simulation, the throughput of a cell surrounded by 18 interfering cells of two tiers was evaluated. Other parameters are summarized in Table 4.2, and reuse factors of $G = 3$ and $G = 4$ were considered, and the respective results are shown in Fig. 4.6 and Fig. 4.7.

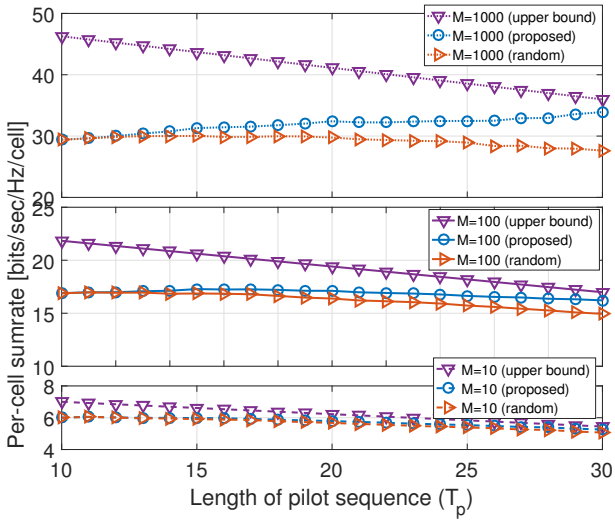

Figure 4.6: Net spectral efficiency per cell with generalized pilot reuse of $G = 3$.

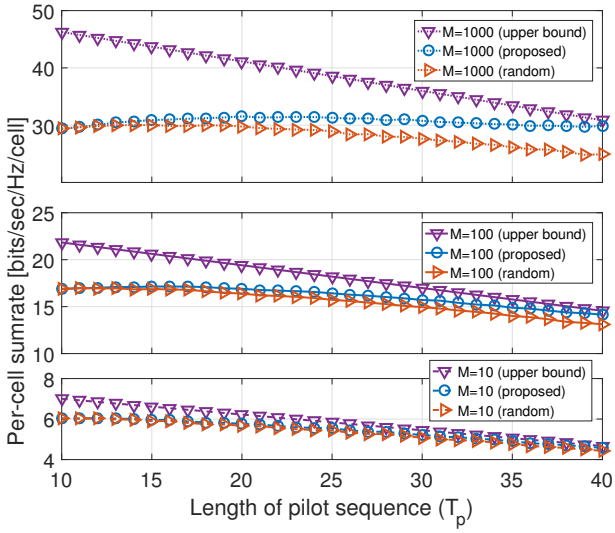

Figure 4.7: Net spectral efficiency per cell with generalized pilot reuse of $G = 4$.

In both cases, we can see that the proposed pilot design, leveraging the Grassmannian subspace packing, outperforms the randomly generated pilot. More importantly,

Table 4.2: Simulation parameters

number of cells	$L = 19$ (18 interfering cells)	cell radius	$r = 1600$ m
number of users per cell	$K = 10$	hole radius	$r_h = 0.2r$
path-loss exponent	$\kappa = 3$	shadow fading deviation	$\sigma_S = 8$ dB
block size	$B = 100$	average SNR	$\rho_u = 10$ dB

the pilot sequences could be designed with an arbitrary length in the generalized pilot reuse framework, while only two corner points are achievable with the conventional pilot reuse. Hence, similarly to the previous section, we can flexibly design pilot sequences for large cellular networks and can achieve a higher spectral efficiency.

4.6 Conclusion

In this chapter, the pilot contamination problem in multi-cell massive MIMO network was investigated. To resolve the pilot contamination, I proposed two approaches. For a small network, an iterative pilot design algorithm was proposed to minimize the pilot contamination, and the generalized pilot reuse framework, along with the Grassmannain training, was proposed for large cellular networks. The proposed methods alleviates pilot contamination in multi-cell systems by minimizing the correlation between the pilot sequences used by different cells. As a result, pilot sequences with an arbitrary length can be designed, hence the spectral efficiency can be improved by adjusting the length of the pilot sequence.

Chapter 5

Conclusion

5.1 Summary

In this dissertation, I developed channel state information (CSI) acquisition techniques for massive multiple-input multiple-output (MIMO) systems. Since systems with different operation modes confront different challenges in acquiring CSI, I proposed a set of solutions that can address challenges for various scenarios. For frequency division duplexing (FDD) systems, I developed a novel channel estimation scheme and a feedback protocol that enable a massive MIMO beamforming with reduced overhead. For time division duplexing (TDD) systems, two pilot design strategies were developed to mitigate the pilot contamination, which is a well-known performance bottleneck in massive MIMO systems. The developed schemes can contribute to the practical use of massive MIMO techniques in future wireless systems.

- To reduce the excessive downlink training overhead in FDD massive MIMO systems, I proposed a compressed sensing (CS)-aided training technique. The proposed scheme exploits partial support information, which can be obtained from previous estimate, for both training sequence design and estimation processing.

Based on the assumption of slowly-varying support, the proposed scheme estimates the channel using the conventional least squares scheme and a CS tool simultaneously. Simulations results showed that the proposed scheme can reduce the training overhead significantly even when the channel is not sparse enough, and hence conventional CS algorithms are not applicable.

- I also proposed a projection-based differential feedback (PBDF) protocol to address problems in massive MIMO channel feedback. Considering that the large dimension of massive MIMO channels poses challenges in feedback in terms of overhead and encoding complexity, I developed a feedback scheme, which projects the differential channel vector into a smaller dimensional space. For the proposed framework, the sequence of optimal projection matrices was also derived incorporating the channel statistics and the codebook structure. The simulation results showed that the proposed scheme achieves performance comparable to ideal massive MIMO systems with a small amount of feedback.

- To address pilot contamination, which causes performance bottleneck in multi-cell massive MIMO systems, I proposed two approaches for uplink channel training. To this end, the effect of the pilot contamination was analyzed, and appropriate design criteria were formulated. Based on these preliminaries, I developed two design strategies. First, a pilot design algorithm based on alternating minimization was developed. Second, a scalable channel training framework, referred to as generalized pilot reuse, was proposed, and it was shown that the Grassmannian subspace packing provides an optimal sequence set for the framework. Simulation results showed that more flexible pilot design becomes viable and a higher throughput can be achieved with the proposed training strategies.

5.2 Future Directions

In this section, I present some future research directions in areas relevant to the CSI acquisition for massive MIMO systems.

- **Advanced downlink training exploiting partial support information:** The downlink training scheme developed in this dissertation exploits partial support information to reduce the training overhead and improve the estimation performance. While the simulation results demonstrated the value of the partial support information, the pilot structure and estimation process of the proposed scheme are naive, and the analysis is far from completeness. As pointed out in Chapter 2, several CS recovery algorithms that utilize partial support information have recently been developed, and their performance was also analyzed [38, 50–53], while the adaptation of sensing matrix was firstly proposed in my work. Developing more advanced adaptation strategy and corresponding recovery algorithm would improve the estimation performance. Analyzing the estimation performance is another interesting direction, which can provide an insight into the gain and application range of the proposed scheme.

- **Analysis of optimal training parameters:** In Chapter 4, it was shown that higher sum-rate can be achieved through more flexible pilot design. However, the optimal length of pilot was obtained only through simulations. From a system design perspective, deriving the optimal length of pilot and the optimal number of users has important meaning. To this end, non-asymptotic behavior (i.e., finite number of antennas) and the performance in random networks need to be investigated. For the non-asymptotic analysis, mathematical tools used in the literature [6, 90] can be utilized. Stochastic geometry can provide a powerful tool for the analysis of random networks [93]. While various random networks in-

cluding cellular networks have been successfully analyzed using this tool [94], and massive MIMO systems were also analyzed from this perspective [95], analysis of pilot reuse system is still open. Although it is a challenging task, analyzing random cellular networks adopting generalized pilot reuse would be a path to complete optimization of massive MIMO systems.

- **Channel estimation for millimeter wave systems:** Developing CSI acquisition techniques for millimeter wave (mmWave) systems is another promising direction. Communications over mmWave bands (30-300 GHz) are drawing significant attention in building future wireless systems. Due to huge spectrum available in those bands, high data rate up to a few Gbps can be achieved in wireless systems operating in mmWave bands [32]. To mitigate high path loss observed in mmWave bands, highly directional beamforming using a large number of antennas is essentially required. The schemes proposed in this dissertation can be utilized also for mmWave systems. However, mmWave exhibits distinct characteristics such as sparse scattering and high path loss, and such characteristics should be considered when developing low-overhead beam alignment and channel estimation techniques [33–35].

  Another challenge for mmWave channel estimation is caused by system architectures different from massive MIMO systems operating in conventional frequency bands. In mmWave systems, due to huge bandwidth and a large number of antennas, conventional beamforming solutions, which require a high-resolution RF chain per antenna, becomes inefficient in terms of cost and power consumption. As potential solutions for low-cost and low-power beamforming, hybrid precoding and the use of low-resolution analog-digital converters (ADCs) are considered [96]. With these architectures, the receiver obtains a limited number of observations or highly quantized signal, and the estimation

becomes more challenging [97]. Developing channel estimation techniques for those low-power, low-cost architecture would be an interesting research direction.

Bibliography

[1] F. Boccardi, R. W. Heath Jr., A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74-80, Feb. 2014.

[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and C. Zhang, "What will 5G be?," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065-1082, Jun. 2014.

[3] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming, *3G Evolution HSPA and LTE for Mobile Broadband*, Elsevier Publishers, Second Edition, 2008.

[4] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40-60, Jan. 2013.

[5] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590-3600, Nov. 2010.

[6] J. Hoydis, S. T. Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160-171, Feb. 2013.

[7] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172-179, Feb. 2013.

[8] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436-1449, Apr. 2013.

[9] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742-758, Oct. 2014.

[10] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112-7139, Nov. 2014.

[11] S. K. Mohammed and E. G. Larsson, "Per-antenna constant envelope precoding for large multi-user MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 3, pp. 1059-1071, Mar. 2013.

[12] C. Risi, D. Persson, and E. G. Larsson, "Massive MIMO with 1-bit ADC," *arXiv*:1404.7736, 2014.

[13] M. Wu, B. Yin, G. Wang, C. Dick, J. R. Cavallaro, and C. Studer, "Large-scale MIMO detection for 3GPP LTE: Algorithm and FPGA implementation," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 916-929, Oct. 2014.

[14] C. K. Au-Yeung, D. J. Love, and S. Sanayei, "Trellis coded line packing: Large dimensional beamforming vector quantization and feedback transmission," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1844-1853, Jun. 2011.

[15] J. Choi, Z. Chance, D. J. Love, and U. Madhow, "Noncoherent trellis coded quantization: A practical limited feedback technique for massive MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 5016-5029, Dec. 2013.

[16] J. Choi, D. J. Love, and T. Kim, "Trellis-extended codebooks and successive phase adjustment: A path from LTE-Advanced to FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2007-2016, Apr. 2015.

[17] Y. Han, W. Shin, and J. Lee, "Projection-based differential feedback for FDD massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 202-212, Jan. 2017.

[18] A. Adhikary, J. Nam, J. Ahn, and G. Caire, "Joint spatial division and multiplexing-The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441-6463, Oct. 2013.

[19] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, "Pilot beam pattern design for channel estimation in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 787-801, Oct. 2014.

[20] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 802-814, Oct. 2014.

[21] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21-30, Mar. 2008.

[22] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principle: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489-509, Feb. 2006.

[23] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207-1223, Aug. 2006.

[24] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030-1051, Mar. 2006.

[25] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 164-174, Nov. 2010.

[26] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058-1076, Jun. 2010.

[27] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563-2579, Oct. 2002.

[28] Y. Zhou, M. Herdin, A. M. Sayeed, and E. Bonek, "Experimental study of MIMO channel statistics and capacity via the virtual channel representation," Univ. Wisconsin-Madison, Tech. Rep., Feb. 2007.

[29] S. L. H. Nguyen and A. Ghrayeb, "Compressive sensing-based channel estimation for massive multiuser MIMO systems," in *Proc. IEEE Wireless Commun. and Netw. Conf. (WCNC)*, Shanghai, China, Apr. 2013.

[30] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261-3271, Jun. 2014.

[31] P. Kuo, H. T. Kung, and P. Ting, "Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays," in *Proc. IEEE Wireless Commun. and Netw. Conf. (WCNC)*, Paris, France, Apr. 2012.

[32] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366-385, Mar. 2014.

[33] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.,* vol. 8, no. 5, pp. 831-846, Oct. 2014.

[34] T. Kim and D. J. Love, "Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels," in *Proc. IEEE Int. Wksp. Signal Process. Adv. Wireless Commun. (SPAWC)*, Stockholm, Sweden, Jun. 2015.

[35] Y. Han and J. Lee, "Two-stage compressed sensing for millimeter wave channel estimation," in *Proc. IEEE Int. Symp. Inf. Theory,* Barcelona, Spain, Jul. 2016.

[36] S. Wu, C. Wang, H. Haas, el-H. M. Aggoune, M. M. Alwakeel, and B. Ai, "A non-stationary wideband channel model for massive MIMO communication systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1434-1446, Mar. 2015.

[37] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication,* Cambridge Univ. Press, 2005.

[38] X. Rao and V. K. N. Lau, "Compressive sensing with prior support quality information and application to massive MIMO channel estimation with temporal correlation," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4914-4924, Sep. 2015.

[39] J. A. Tropp and C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.

[40] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680-4688, Jul. 2011.

[41] R. Wu and D. Chen, "The exact support recovery of sparse signals with noise via orthogonal matching pursuit," *IEEE Signal Process. Lett.*, vol. 20, no. 4, pp. 403-406, Apr. 2013.

[42] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203-4215, Dec. 2005.

[43] A. S. Bandeira, E. Dobriban, D. G. Mixon, and W. F. Sawin, "Certifying the restricted isometry property is hard," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3448-3450, Jun. 2013.

[44] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253-263, Dec. 2008.

[45] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695-5702, Dec. 2007.

[46] G. Li, Z. Zhu, D. Yang, L. Chang, and H. Bai, "On projection matrix optimization for compressive sensing systems," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2887-2898, Nov. 2013.

[47] P. Kyritsi, D. C. Cox, R. A. Valenzuela, and P. W. Wolniansky, "Correlation analysis based on MIMO channel measurements in an indoor environment," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 713-720, Jun. 2003.

[48] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477-2488, Jul. 2005.

[49] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572-588, Mar. 2006.

[50] N. Vaswani, "LS-CS-residual (LS-CS): Compressive sensing on least squares residual," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4108-4120, Aug. 2010.

[51] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4595-4607, Sep. 2010.

[52] M. P. Friedlander, H. Mansour, R. Saab, and Ö. Yilmaz, "Recovering compressively sampled signals using partial support information," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1122-1134, Feb. 2012.

[53] J. Ziniel and P. Schniter, "Dynamic compressive sensing of time-varying signals via approximate message passing," *IEEE Signal Process.*, vol. 61, no. 21, pp. 5270-5284, Nov. 2013.

[54] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, pp. 18914-18919, Nov. 2009.

[55] J. Shen, J. Zhang, E. Alsusa, and B. Letaief, "Compressed CSI acquisition in FDD massive MIMO: How much training is needed?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4145-4156, Jun. 2016.

[56] D. Ying, F. W. Vook, T. A. Thomas, D. J. Love, and A. Ghosh, "Kronecker product correlation model and limited feedback codebook design in a 3D channel model," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Sydney, Australia, Jun. 2014.

[57] D. J. Love, R. W. Heath Jr., V. K. N. Lau, D. Gesberts, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341-1365, Oct. 2008.

[58] D. J. Love, R. W. Heath Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2735-2747, Oct. 2003.

[59] C. K. Au-Yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in MISO systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 458-462, Feb. 2007.

[60] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045-5060, Nov. 2006.

[61] K. Ko, S. Jung, and J. Lee, "Hybrid MU-MISO scheduling with limited feedback using hierarchical codebooks," *IEEE Trans. Commun.*, vol. 60, no. 4, pp. 1101-1113, Apr. 2012.

[62] D. J. Love and R. W. Heath Jr., "Limited feedback diversity techniques for correlated channels," *IEEE Trans. Veh. Technol.*, vol. 55, no. 2, pp. 718-722, Mar. 2006.

[63] V. Raghavan, R. W. Heath Jr., and A. M. Sayeed, "Systematic codebook designs for quantized beamforming in correlated MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1298-1310, Sep. 2007.

[64] T. Kim, D. J. Love, and B. Clerckx, "MIMO systems with limited rate differential feedback in slowly varying channels," *IEEE Trans. Commun.*, vol. 59, no. 4, pp. 1175-1189, Apr. 2011.

[65] J. Choi, B. Clerckx, N. Lee, and G. Kim, "A new design of polar-cap differential codebook for temporally/spatially correlated MISO channels," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 703-711, Feb. 2012.

[66] J. Mirza, P. Dmochowski, P. Smith, and M. Shafi, "Limited feedback multiuser MISO systems with differential codebooks in correlated channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, Jun. 2013.

[67] Y. Sun, J. Zhang, P. Zhang, L. Wu, and Y. Wang, "Practical differential quantization for spatially and temporally correlated massive MISO channels," in *Proc. IEEE Int. Symp. Personal, Indoor, and Mobile Radio Commun. (PIMRC)*, Washington, DC, Sep. 2014.

[68] B. Lee, J. Choi, J. Seol, D. j. Love, and B. Shim, "Antenna grouping based feedback compression for FDD-based massive MIMO systems," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3261-3274, Sep. 2015.

[69] Z. Zhang, K. C. Teh, and K. H. Li "Application of compressive sensing to limited feedback strategy in large-scale multiple-input single-output cellular networks," *IET Commun.*, vol. 8, no. 6, pp. 947-955, Apr. 2014.

[70] J. Nam, A. Adhikary, J. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 876-890, Oct. 2014.

[71] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, "Achievable rates of FDD massive MIMO systems with spatial channel correlation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2868-2882, May 2015.

[72] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528-541, Mar. 2006.

[73] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951-963, Apr. 2003.

[74] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression,* Kluwer Academic Publishers, 1991.

[75] J. G. Proakis and M. Salehi, *Communication Systems Engineering,* Prentice Hall, 2002.

[76] I. Viering, H. Hofstetter, and W. Utschick, "Validity of spatial covariance matrices over time and frequency," in *Proc. IEEE Global Telecom. Conf. (Globecom)*, Taipei, Taiwan, Nov. 2002.

[77] G. Xu and H. Liu, "An effective transmission beamforming scheme for frequency-division-duplex digital wireless communication systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Detroit, MI, May 1995.

[78] G. C. Raleigh, S. N. Diggavi, V. K. Jones, and A. Paulraj, "A blind adaptive transmit antenna algorithm for wireless communication," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Seattle, WA, Jun. 1995.

[79] Y. C. Liang and F. P. S. Chin, "Downlink channel covariance matrix (DCCM) estimation and its applications in wireless DS-CDMA systems," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 2, pp. 222-232, Feb. 2001.

[80] M. K. Tsatsanis, G. B. Giannakis, and G. Zhou, "Estimation and equalization of fading channels with random coefficients," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Atlanta, GA, May 1996.

[81] K. Han, S. Lee, J. Lim, and K. Sung, "Channel estimation for OFDM with fast fading channels by modified Kalman filter," *IEEE Trans. Consum. Electron.*, vol. 50, no. 2, pp. 443-449, May 2004.

[82] G. Barriac and U. Madhow, "Space-time communication for OFDM with implicit channel feedback," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3111-3129, Dec. 2004.

[83] D. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502-513, Mar. 2000.

[84] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, "Inter-cell interference in noncooperative TDD large scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 192-201, Feb. 2013.

[85] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264-273, Feb. 2013.

[86] Y. Li, Y.-H. Nam, B. L. Ng, and J. C. Zhang, "A non-asymptotic throughput for massive MIMO cellular uplink with pilot reuse," in *Proc. IEEE Global Telecom. Conf. (Globecom)*, Anaheim, CA, Dec. 2012.

[87] I. Atzeni, J. Arnau, and M. Debbah, "Fractional pilot reuse in massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC) Workshop*, London, UK, Jun. 2015.

[88] R. Mochaourab, E. Björnson, and M. Bengtsson, "Pilot clustering in asymmetric massive MIMO networks," in *Proc. IEEE Int. Wksp. Signal Process. Adv. Wireless Commun. (SPAWC)*, Stockholm, Sweden, Jun. 2015.

[89] H. Wang, W. Zhang, Y. Liu, Q. Xu, and P. Pan, "On design of non-orthogonal pilot signals for a multi-cell massive MIMO system," *IEEE Wireless Commun. Lett.*, vol. 4, no. 2, pp. 129-132, Apr. 2015.

[90] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1293-1308, Feb. 2016.

[91] I. Csiszár and G. Tusnády, "Information geometry and alternating minimization procedures," *Statistics & Decisions*, Supplement Issue, no. 1, pp. 205-237, 1984.

[92] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: Packings in Grassmannian spaces," *Exper. Math.*, vol. 5, no. 2, pp. 139-159, 1996.

[93] F. Baccelli and B. Błaszczyszyn, *Stochastic Geometry and Wireless Networks.* NOW: Foundations and Trends in Networking, 2010.

[94] H. ElSawy, A. Sultan-Salem, M.-S. Alouini, and M. Z. Win, "Modeling and analysis of cellular networks using stochastic geometry: A tutorial," *IEEE Commun. Surveys Tuts.*, preprint.

[95] T. Bai and R. W. Heath Jr., "Analyzing uplink SINR and rate in massive MIMO systems using stochastic geometry," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4592 - 4606, Nov. 2016.

[96] A. Alkhateeb, J. Mo, N. González-Prelcic, and R. W. Heath Jr., "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122-131, Dec. 2014.

[97] J. Mo, P. Schniter, and R. W. Heath Jr., "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *arXiv*:1610.02735, 2016.

초 록

지속적으로 증가하는 무선 통신 수요에 대응하기 위해 기존 시스템에 비해 전송 효율을 획기적으로 높일 수 있는 차세대 통신 시스템 설계에 대한 관심이 높아지고 있다. 이를 위한 다양한 후보 기술 중 massive MIMO(large-scale antenna 또는 대규모 다중 안테나) 기술은 기존의 다중안테나 기술의 공간 다중화 이득을 극대화하기 위해 수백 개 이상의 매우 많은 안테나를 활용하는 기술로, 이를 적용한 시스템에서는 큰 수의 법칙에 의해 설명되는 서로 다른 이용자 채널 사이의 직교성 덕분에 단순한 신호처리 만으로도 높은 전송률을 얻을 수 있다. 하지만, 이러한 이득을 얻기 위해서는 정확한 채널 정보를 활용하는 것이 필수적이며 안테나 수에 비례해 채널의 차원이 커지기 때문에 채널 정보 획득에 어려움이 발생한다. 하향링크 트레이닝과 상향링크 피드백을 채널 정보를 얻는 주파수 분할 이중(frequency division duplexing; FDD) 시스템에서는 채널 트레이닝과 피드백을 위한 오버헤드가 안테나의 수에 비례해 증가하기 때문에 안테나 수가 매우 많은 massive MIMO 시스템에서는 오버헤드를 고려한 전송 효율이 매우 낮아지게 된다. 상향링크-하향링크 채널이 동일한(reciprocal) 특성을 갖는 시분할 이중(time division duplexing; TDD) 시스템에서는 상향링크 트레이닝을 통해 하향링크 채널 정보 또한 얻을 수 있지만, 상관된 파일럿의 사용으로 인해 발생하는 pilot contamination은 안테나 수의 증가로 해결할 수 없는 근본적인 문제로 지적되어왔다.

본 논문에서는 다양한 형태의 massive MIMO 시스템을 위한 효율적인 채널 정보 획득 방법을 제시한다. 먼저, FDD 시스템에서 적은 채널 이용으로 하향링크 채널을 추정하기 위한 트레이닝 기법을 제안한다. 이를 위해 압축 센싱 기법을 이용하며 이전 채널 정보를 적절히 활용함으로써 오버헤드를 크게 감소시킬 수 있음을 확인한다. 또, FDD 시스템의 채널 피드백 오버헤드를 줄이기 위한 제한적 피드백 기법을 제시한다. 제안하는 기법은 채널의 시간적/공간적 상관관계를 활용한 차원 감소 기

법을 활용해 피드백에 필요한 오버헤드를 크게 감소시킬 수 있다. 마지막으로, 다중 셀 massive MIMO 시스템에서 근본적인 한계로 지적되는 pilot contamination의 영향을 분석하고, 이를 경감하기 위한 두 가지 상향링크 파일럿 설계 방식을 제안한다. 주어진 네트워크에서 최적의 파일럿을 설계하는 반복적 설계 기법과, 셀의 수가 많은 네트워크에 활용할 수 있는 generalized pilot reuse 기법을 제시한다.

주요어: massive MIMO, 대규모 다중 안테나 시스템, 채널 추정, 제한적 피드백, pilot contamination

학번: 2011-20954